RESEARCH ARTICLE

A method to estimate prey density from single-camera images: A case study with chinstrap penguins and Antarctic krill

Victoria R. Hermanson[1¤], George R. Cutter[1], Jefferson T. Hinke[1]*, Matthew Dawkins[2], George M. Watters[1]

1 Antarctic Ecosystem Research Division, Southwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, La Jolla, CA, United States of America, 2 Kitware Inc., Clifton Park, NY, United States of America

¤ Current address: University of California Santa Cruz, Santa Cruz, CA, United States of America
* Jefferson.Hinke@noaa.gov

Abstract

Estimating the densities of marine prey observed in animal-borne video loggers when encountered by foraging predators represents an important challenge for understanding predator-prey interactions in the marine environment. We used video images collected during the foraging trip of one chinstrap penguin (*Pygoscelis antarcticus*) from Cape Shirreff, Livingston Island, Antarctica to develop a novel approach for estimating the density of Antarctic krill (*Euphausia superba*) encountered during foraging activities. Using the open-source Video and Image Analytics for a Marine Environment (VIAME), we trained a neural network model to identify video frames containing krill. Our image classifier has an overall accuracy of 73%, with a positive predictive value of 83% for prediction of frames containing krill. We then developed a method to estimate the volume of water imaged, thus the density ($N \cdot m^{-3}$) of krill, in the 2-dimensional images. The method is based on the maximum range from the camera where krill remain visibly resolvable and assumes that mean krill length is known, and that the distribution of orientation angles of krill is uniform. From 1,932 images identified as containing krill, we manually identified a subset of 124 images from across the video record that contained resolvable and unresolvable krill necessary to estimate the resolvable range and imaged volume for the video sensor. Krill swarm density encountered by the penguins ranged from 2 to 307 krill·m^{-3} and mean density of krill was 48 krill·m^{-3} (sd = 61 krill·m^{-3}). Mean krill biomass density was 25 g·m^{-3}. Our frame-level image classifier model and krill density estimation method provide a new approach to efficiently process video-logger data and estimate krill density from 2D imagery, providing key information on prey aggregations that may affect predator foraging performance. The approach should be directly applicable to other marine predators feeding on aggregations of prey.

Introduction

The use of animal-borne instrumentation has expanded studies on the physiology [1] and ecology [2] of free-ranging marine predators. A variety of telemetry devices have enabled the detailed characterization of diving energetics and behaviors [3, 4], movement and migration patterns [5, 6], habitat partitioning [7, 8], and identification of essential marine habitats [9, 10]. Accelerometers mounted on the head or jaw [11–13], ingested stomach temperature probes [14], and movement data [15], (but see [16]) have been used to estimate feeding rates. Direct observation of predation events, prey, and foraging habitats have been done for large animals such as seals [17–20] and whales [21–25], but assessing predator responses to changes in prey availability remains difficult in the marine environment, particularly for smaller animals such as penguins [26–30]. While there are studies that have observed penguins striking at prey [12, 31, 32], in this work we quantify the prey field and prey availability, which may influence feeding behaviors and success.

Miniaturization of digital cameras has allowed first-person views of underwater behaviors in a suite of marine species [33–35], including penguins [23, 27, 36–39]. Importantly, these image-based methods offer direct observations of the interactions between predator and prey and provide the potential for novel inference on key ecological interactions. Images and data derived from animal-borne video loggers may provide a useful step toward understanding predator responses to variation in prey density. For example, video data allow assessments of the frequency of encounters with prey and estimates of predator foraging rate within prey patches [40, 41]. However, estimating the density of prey from two-dimensional (2D) images obtained from animal-borne cameras represents a significant challenge for two main reasons. First, visual analysis and manual annotation of video imagery to identify prey encounters is slow and inefficient, requires subjective interpretation, and may be affected by observer error and bias. Second, when prey are visible in imagery, counts of individuals within the image can provide a relative index of prey abundance, but estimating the concentration of prey (volumetric density) additionally requires estimating the imaged volume. To overcome these challenges, we trained an automated detection algorithm to classify video content at the frame-level and developed a method to estimate the volumetric density observed in 2D imagery of prey encountered by a foraging predator.

Machine learning algorithms are increasingly used in the ecological studies to automate detection of features of interest in imagery [42, 43]. Frame-level classifier models can be trained to identify images that contain characteristics indicative of behavior, locations, or the environment. For example, in animal-borne video data from an air-breathing marine predator, the characteristics of images obtained at the surface are distinct from those obtained underwater, and images obtained during open-water diving are different from those obtained during encounters with prey aggregations. Classifier models can be trained to differentiate such characteristics [44, 45] and identify frames with prey encounters for further analysis. We leverage automated image processing to alleviate the burden of time spent manually processing hours of video imagery. The open-source computer vision platform VIAME (Video and Image Analytics for the Marine Environment [46]; available from https://www.viametoolkit.org/) provides a flexible and powerful tool for image annotation, detector training, and image analysis. We used VIAME to develop a classifier model to automate identification of images containing krill based on full frame-level content.

Several approaches can be used to estimate three-dimensional (3D) scene structure from single camera 2D imagery. An approach developed to remove haze from images, to estimate light transmission, and enhance underwater imagery using the dark channel prior [47, 48] produces a depth map of the scene that is useful for characterizing topography and relative distances among scene elements. Most methods for 3D estimation from single images produce
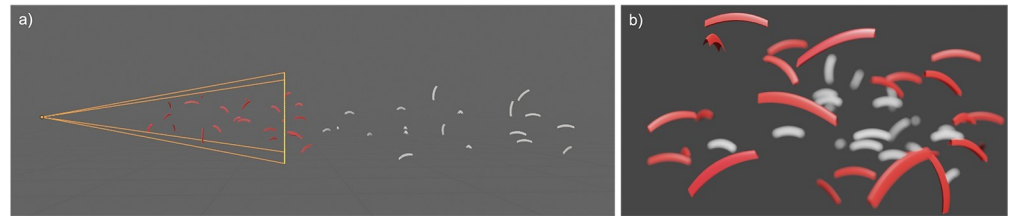
Fig 1. Two-panel view of the resolvable range of the camera. a) side view and b) camera view of targets with simulated depth of field effects.

depth maps [49] with relative distances to objects that must be scaled based on known sizes of scene elements at known ranges to estimate actual coordinates or sizes.

Other single-camera, single image approaches for estimating distance to imaged targets use object focus methods, also known as the depth from defocus [50], to reconstruct 3D scenes from individual 2D images without the assistance of additional sensors or multiple cameras [51–54]. Depth-from-defocus methods assume that imaged targets within a 2D scene can be unresolved due to large ranges from the imaging system, limits of resolution, or because of relative motion of the object and camera. Assuming that targets have a known length, combined with a distinction between resolvable (containing discernible visible features) and unresolvable targets (lacking distinctive features or out of focus because of range and not as a result of motion blur) in the image can enable the estimation of the imaged range, therefore reconstructing a 3D scene (Fig 1).

Here, we used animal-borne video imagery to first train a frame-level image classifier with a convolutional neural network model to identify images that represent basic penguin behaviors including surfacing/diving, surface resting, underwater transit/swimming, and prey encounters. Based on a subset of images containing krill, we then implement a method conceptually similar to the approaches for estimating single-image depth of field and depth from defocus [53–55] to estimate the volumetric density ($N \cdot m^{-3}$) of krill swarms encountered by foraging chinstrap penguins.

Materials and methods

Instrumentation

We collected video imagery of the interactions between chinstrap penguins and their principal prey, Antarctic krill, from Cape Shirreff, Livingston Island (60.79˚W, 62.46˚S) in the northern Antarctic Peninsula region. Videos were collected in February 2018 and December 2019 using Little Leonardo DVL400M028 (52 mm × 20 mm ×11mm, 15 g) and DVL400M065 (61 mm × 21 mm × 15 mm, 29 g) digital video loggers (DVLs). The DVLs recorded color imagery using ambient light at 30 frames per second (fps). Field-of-view angles were 31° horizontal by 24° vertical and the image frame size was 1280 by 960 pixels (pix). Videos were recorded continuously for 5 to 8 hours (separated by 30-minute intervals when exported) until memory capacity was full or the battery was exhausted. Each penguin was also instrumented with a time depth recorder (TDR; Lotek LAT1800FP, 36x13x11 mm, 9 g) to record dive depth. Instrument attachment and recovery methods were described previously [40].

Ethical approval

All animal handling procedures and research protocols are approved by the Southwest Fisheries Science Center/Pacific Islands Fisheries Science Institutional Animal Care and Use Committee (# SWPI 2020–01). All field research activities are permitted under the U.S. Antarctic Conservation Act (Permit #2017–012).

Table 1. Number of training and test images for each class. Each class is representing the principle visual content associated with penguin behaviors and observed conditions while foraging. Training and test images were taken from video collected in February 2018 and December 2019.

Training label	Training Images	Test Images
Open water	815	320
Krill presence	147	262
Surface	553	324
Penguin presence	2	21
Bright or dark	121	73

https://doi.org/10.1371/journal.pone.0303633.t001

Automated image analysis

We used the imagery from both 2018 and 2019 to train a frame-level image classifier model. We used VIAME (version 0.19.4) to train an EfficientNet convolutional neural network [56] to classify the principal frame-level content of video images. The model was trained on a 1,638 image subset using a five-class scheme (Table 1). Each of the classes represent the frame-level visual content associated with behaviors or observed conditions recorded during foraging dives that represent: 1) **open water**, indicative of underwater swimming; 2) **surface**, indicative of surface activities such as breathing and resting during intervals between dives; 3) **bright or dark**, very bright or dark conditions indicative of passing through the air-water interface during diving or surfacing events; 4) **penguin presence** (while swimming underwater) indicative of group association; 5) **krill presence**, indicative of prey encounters (Fig 2). The number of images per class in the training set ranged from 2 to 815 (Table 1).
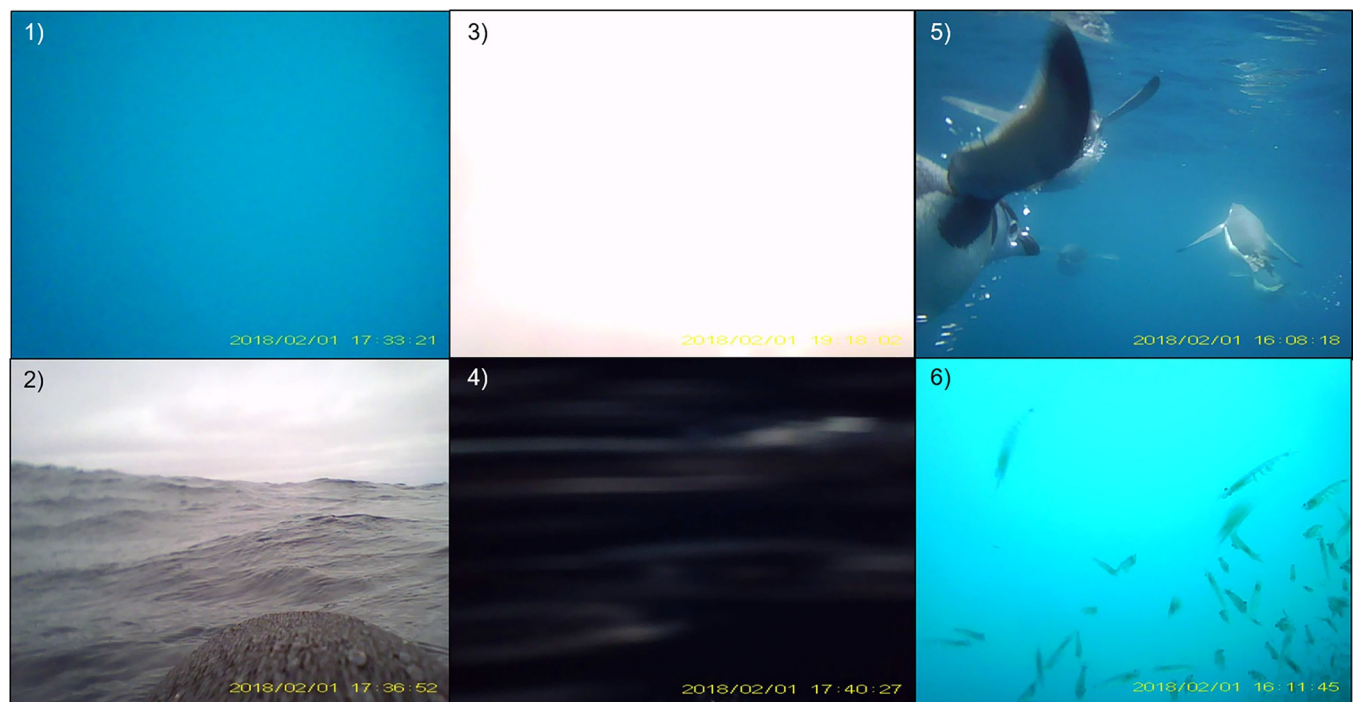

Fig 2. Example images of the frame-level classes used for model training. The five classes represent images from 1) open water representing underwater transit/swimming, 2) surface, representing resting on the surface, 3) bright, during surfacing or diving, 4) dark, during surfacing or diving, 5) penguin presence, 6) krill presence.

https://doi.org/10.1371/journal.pone.0303633.g002

Model performance evaluation

Model performance was evaluated with a confusion matrix using the R (R Core Team 2020) package 'caret' [57]. The trained model was tested on a subset of 1,000 new images. In conjunction, manual annotations (a truth set) were completed on the test subset to use for the confusion matrix evaluation.

Confusion matrix metrics relating true classes to model-predicted classes were represented using counts of true positives, false positives, true negatives, and false negatives from model predictions. True Positives (TP) represent correct model predicted class matching the known, analyst-designated ("truth") class of an image. True Negatives (TN) are all images that the model correctly identifies as not containing the class; False Positives (FP) are cases where the predicted class does not represent the true class. False Negatives (FN) are images predicted as negative for a class when they should represent the true class. For each class we calculated:

$$accuracy: \ A = \frac{TP + TN}{TP + FP + FN},$$

$$precision \ (positive \ predicted \ value): \ P = \frac{TP}{TP + FP},$$

$$recall \ (sensitivity): \ R = \frac{TP}{TP + FN},$$

$$specificity = \frac{TN}{TN + FP},$$

$$\begin{aligned}negtive \ predictive \ value: \ NPV \\ = (specificity*(1 - prevalence))/(((1 - sensitivity)*prevalence) + ((specificity*(1 - prevalence)),\end{aligned}$$

$$balanced \ accuracy = \frac{sensitivity + specificity}{2},$$

$$prevalence = \frac{N_{class}}{N_{total}}, \ where \ N_{total} \ is \ the \ count \ by \ class \ and \ N_{total} \ is \ the \ total \ number \ of \ images \ in \ the \ truth \ set, \ and \ the$$

$$F1 \ score: \ F1 = \frac{2*(P*R)}{P + R}.$$

[58–60]

These metrics were calculated to evaluate the performance of the classifier model. After model performance was evaluated, our classifier model was used to detect frames containing from a new image set (1.5 hours of video sampled at 5 fps collected in February 2018) to be used for prey density estimation.

Estimating prey density from 2D imagery

The images used to estimate krill density were manually selected from a set of images that the EfficientNet classifier model predicted to be krill from 1.5 hours of video sampled at 5 fps collected in February 2018. From these we selected a subset of 124 images where motion blur was minimal and the image included targets distributed from near to far, thus allowing classification of resolvable and unresolvable krill (Fig 1).
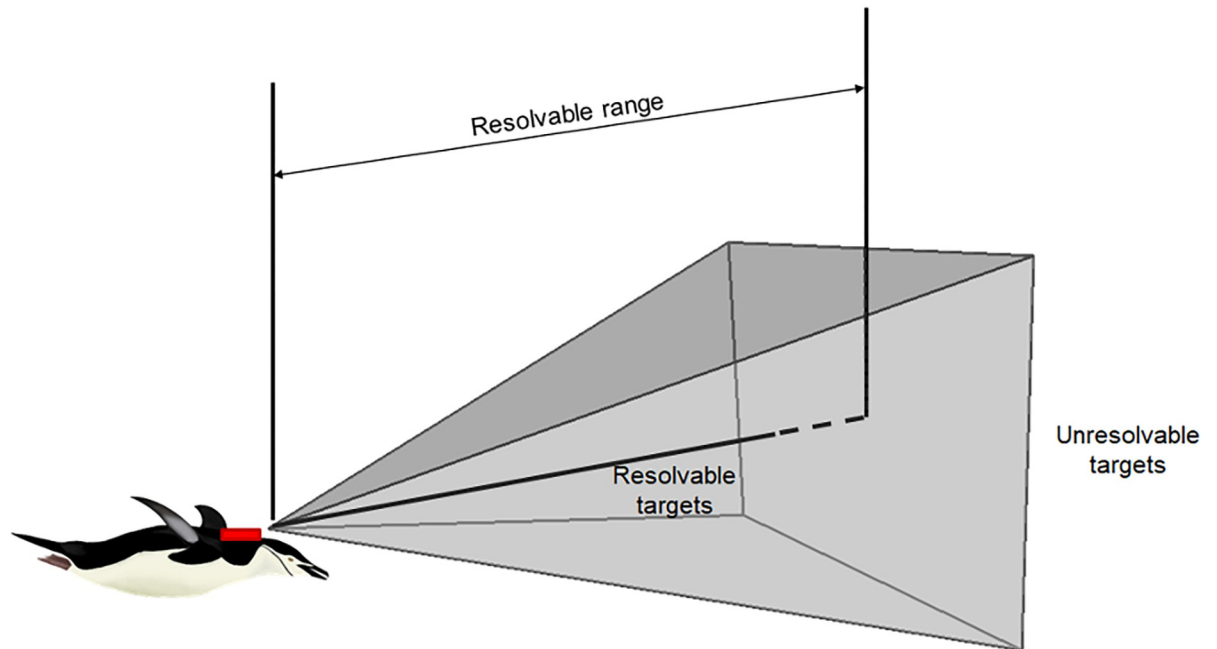
Fig 3. Conceptual image for estimating the resolvable range of the camera and the imaged volume. Only resolvable targets are included within the resolvable range. Imaged volume was estimated from the resolvable range (r_{res}) and the field-of-view of the camera lens.

https://doi.org/10.1371/journal.pone.0303633.g003

The density of krill in an image was estimated by assuming the imaged volume of a photograph to be represented by a triangular pyramid (Fig 3). We define the "resolvable range" (r_{res}) as the maximum range within which imaged targets (krill) are clearly identifiable. By relating the imaged length of targets at the limit of the resolvable range to known krill lengths, and using information on the dimensions of the image, we estimated the physical range to these targets to solve for the imaged volume. Counts of resolvable krill within the imaged volume were then converted to density estimates. Camera and lens calibrations were not done, but distortions were negligible based on visual inspection of images. For example, objects like other penguins had consistent appearances in various locations (e.g. center vs. corner) and the horizon appeared as a straight line in images from surface intervals.

Resolvable range and imaged volume

To estimate the densities of krill encountered during a penguin's foraging event, we first estimate the volume of water sampled by the images by estimating the resolvable range (Fig 3). Estimating resolvable range minimally requires an estimate of the length of krill (in pixels) at the distance representing the limit of the imaged volume and an estimate of the true length (in mm) of the imaged targets.

Krill imaged length at the resolvable range limit was estimated by manually classifying all imaged targets as either resolvable or unresolvable (Fig 4). For resolvable krill, we also annotated whether they presented as axial (when a krill is perpendicular to the camera), partial (when part of the animal was out of the frame), bent, or motion-blurred. Resolvable krill were identified based on distinguishing features such as segmentation of the thoracic region, eyes, carapace, telson, or uropods. In some instances, krill body parts are not visible, but the body shape is distinguishable, and in general this was the strongest indicator that an object was a krill. Unresolvable objects are visually indistinguishable and lack any landmark features or

Fig 4. Example of resolvable annotations in an image. Determining the resolvability of krill is dependent upon the contents within each individual frame. Krill within the resolvable range of the camera are apparent; however, smaller targets presumed to be krill are classified as an unresolvable swarm because the distinguishing features of krill (eyes, appendages, body shape) are not visibly distinct.

shape attributes that might reliably indicate krill even in the context of a monospecific swarm aggregation.

Note that body orientation will affect measured length in the images. We observed krill in any orientation without apparent tendency for a particular direction relative to the penguin or camera, where krill orientations range from full normal incidence (side, dorsal, or ventral) with projected imaged size as complete length, to axial (head or tail) with projected size representing width. We accounted for this variation in body orientation relative to the camera by assuming that krill present in a uniform distribution of orientation angles from -180 to 180 (Fig 1). As such, the median of imaged lengths projected through a uniform distribution of angles, accounting for non-zero axial length, was 0.707 times the full normal incidence length. This is equivalent to a length projected at +-45 or +-135 degrees from normal incidence. The inverse of this value, 1.414, was used to compensate mean and median imaged lengths (not individual lengths) for the orientation distribution.

Because the images are acquired using a single camera, we do not have a direct measurement of range to imaged objects, therefore we estimate the resolvable range based on the size at the transition from resolvable to unresolvable objects. We assume that resolvable and unresolvable objects generally correspond to objects within and beyond the imaged volume,

respectively. Also, because we expect that krill have a narrow size distribution, the size distributions of resolvable and unresolvable objects would be differentiable because of the effect of distance on imaged size (resolvable being closer/large and unresolvable being farther/smaller, in general). The measured image size distributions of unresolvable and resolvable targets are overlapping (Fig 5) and do not indicate a sharp transition or bound on the imaged volume, but they do allow us to estimate the resolvable range.

To estimate resolvable range, we fit a logistic regression model (binary classifier) to imaged sizes of objects assigned to the resolvable (krill) and unresolvable object classes. Size data were excluded for partial krill (along edges), and tightly packed aggregations not individually labeled. The resolvable range was then estimated based on the class division boundary estimated by solving a logistic regression model at p = 0.5 where $L_{boundary}$ = 34.4 pixels. The median unresolvable object size, 39.6 pixels, was not much larger than this, and could perhaps be a suitable alternative and more simple to calculate.

To characterize the error and extent of the resolvable range we assumed that all objects with sizes smaller than the logistic regression class size boundary value were unresolvable and that this set of size values represented half the size distribution of objects close to an indistinct boundary marking the limit of the imaged volume. Therefore, we mirrored and duplicated this set of object sizes to produce a symmetrical distribution of size values representing the boundary region and near-boundary objects. This process began by shifting the size boundary set to have a maximum of zero, then we combined that set of values with the absolute value, where we finally shifted back to the boundary value. We did this instead of using the full unresolvable object size distribution (Fig 5) because the unresolvable class included many objects with larger-than-expected sizes that were believed to be krill located within the imaged volume but affected by motion blur strong enough not to allow identification. By strict adherence to the class definitions, these were assigned as unresolvable, and this appeared to have biased the distribution. We fit a normal distribution to the mirrored near-boundary size data and estimated mean and standard deviation.

Resolvable range in imaged units r_{res} (pixels) is estimated as the distance from the camera based on known image dimensions (width: 1,280 pixels, height: 960 pixels) and the camera lens angles (31 by 24˚). We assumed that the imaged krill were similar in size to krill obtained from penguin diet samples that were collected [61] and measured [62] during the 2018 field season. Stomach samples were obtained from 20 chick-provisioning penguins in January and February of 2018. The lengths of up to 55 (47 +/- 10, mean +/- 1 sd) intact krill per sample were retrieved and measured for total length to the nearest mm. Resolvable range, width, and height (pixels) are then converted to r_{res} (m) physical units using an imaged-to-physical length conversion factor (mm/pixels) based on the median length (mm) of krill from diet samples divided by the near-boundary imaged length of krill. Imaged volume is calculated as $V = (w * h * r_{res})/3 m^3$.

Sensitivity and error

To characterize the sensitivity of our method to uncertainty in identification of a range boundary defining a transition from resolvable to unresolvable targets, we estimated resolvable range (m), imaged volume (m^3), and krill density (N·m^3) at a discrete set of imaged sizes representing the near boundary distribution statistics: the mean size at $L_{boundary}$ +- 1 and 2 standard deviations. To characterize the far limit of the volume, we also estimated range, volume, and density using the minimum imaged size of resolvable krill (14.9 pixels) and unresolvable objects (9.9 pixels). To characterize error, we generated a random sample (n = 1000) from a normal distribution with parameters fit to observed near-boundary sizes ~N(34.4, 10.5 pixels), excluding zero and negative values, and estimated the corresponding distributions of resolvable range, imaged volume, and krill density.
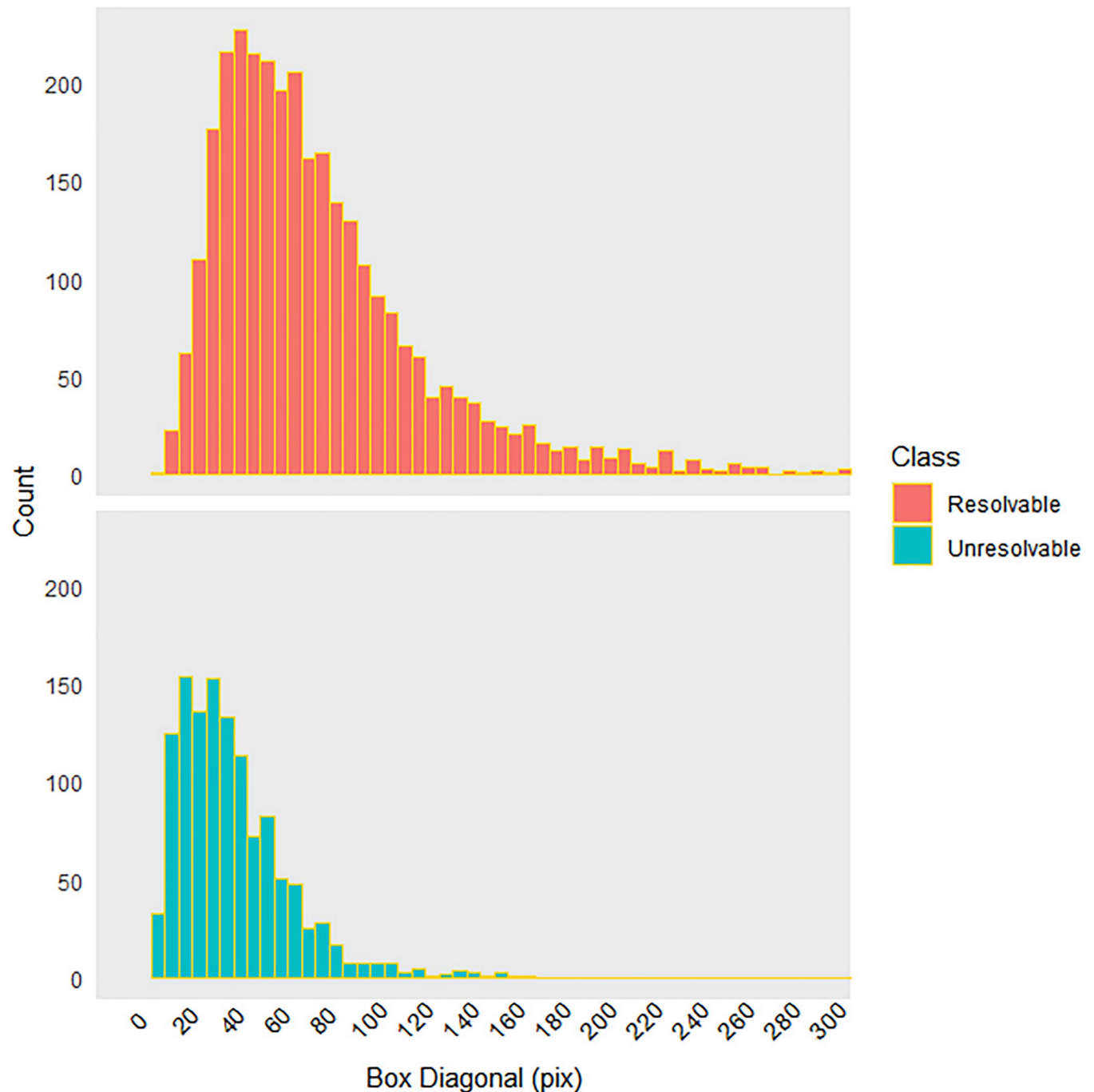
Fig 5. Length frequency distributions (pixel length) for resolvable and unresolvable Antarctic krill (*Euphausia superba*). There is overlap between the lower values of the resolvable krill imaged lengths and the upper values of the unresolvable krill imaged lengths indicating the range of where our resolvable range limit would occur.

https://doi.org/10.1371/journal.pone.0303633.g005

Krill density and biomass

All krill that were classified as resolvable were used to calculate krill numerical density (krill·m^{-3}) for the number of resolvable krill per image, divided by our aforementioned estimated imaged volume. A length-weight relationship for krill in the Antarctic Peninsula region

[62]: $w = aL^b$ (with L as total length in mm), a = 2.236x10^{-6}, and b = 3.314, was used to estimate mean krill weight w = 0.54 g, and allowed conversion between estimated krill densities (krill·m^{-3}) to biomass density (g·m^{-3}).

Comparison of model predictions to manual annotations

A separate data set with manual annotations of the timing of krill strikes, diving, and surfacing based on previously reported methods [40] was compared to model predictions of frame-level image content. From this, we report summaries of foraging behaviors, including dive and feeding durations, and estimates of swarm thickness (i.e. vertical distance in meters) based on the depth range between the start and end of strikes at krill.

Results

Frame-level content classifier

The classifier model was overall 73% accurate in being able to correctly identify frame-level image content. For the krill class performance, metrics indicated an accuracy of 55%, precision of 83% and recall of 63% (Table 2). Performance metrics per class varied. The surface class had particularly high values of 92% accuracy and 99% recall, whereas the penguin presence class had a low value of 15% accuracy and precision. On the other hand, the open water class resulted in 46% accuracy and 67% precision (Table 2).

Krill length and weight

The length of krill measured from diet samples ranged from 31 to 60 mm with a mean length of 43 mm (SD = 4.3 mm) and a median of 42 mm. We assume that these lengths are representative of the physical length distributions of krill observed in images collected in the 2017/2018 field season. Mean weight was estimated as 0.54 g (std. dev = 0.2 g).

Resolvable range and imaged volume

The classification model predicted that 1,932 of the 27,006 frames (7.2%) contained krill. From these predictions, we selected 124 images (all from December 2018) with minimal motion blur for analysis. Manual classification of resolvability yielded 5,055 krill assigned to resolvability classes, 3,079 resolvable, 154 bent, 193 partial, 12 axial, 125 motion-blurred, and 1,474 unresolvable objects. Imaged lengths ranged from 15 to 766 pixels for resolvable krill, and 10 to 177 pixels for unresolvable objects (Fig 5). The median imaged length of resolvable krill was 72 pixels and the mean imaged length was 85 pixels (sd = 55 pixels), with less than 1% of resolvable krill exceeding 400 pixels. The mean imaged length of unresolvable objects was 45 pixels (sd = 25 pixels, median = 40 pixels).

Table 2. Class-specific performance evaluation metrics for a full-frame classification model.

Statistic	Bright-dark	Krill	Surface	Open water	Penguin presence
Accuracy	0.49	0.55	0.92	0.46	0.15
Recall	0.49	0.63	0.99	0.59	0.95
Specificity	0.99	0.95	0.96	0.86	0.89
Precision	0.84	0.83	0.93	0.67	0.15
Negative predicted value	0.96	0.88	1.00	0.82	1.00
Prevalence	0.07	0.26	0.32	0.32	0.02
Balanced Accuracy	0.74	0.79	0.98	0.73	0.92
F-1 score	0.62	0.72	0.96	0.63	0.26

Table 3. Summary statistics for each of the boundary regions. As defined, including the size in pixels at the boundary, the resolvable range from the pixel value, imaged width, imaged height, imaged volume, krill density mean, and krill biomass based off of the density mean.

Boundary	Size at boundary (pix)	Res range (mm)	Imaged width (mm)	Imaged height (mm)	Imaged volume (m³)	Krill density (krill·m⁻³) (mean)	Krill biomass (g·m⁻³)
size_unres_min	9.9	6774	3839	3839	24.96	1.15	3.86
size_res_min	14.9	4511	2557	2557	7.37	3.90	13.09
sizeatclassdivision_mean	**34.4**	**1947**	**1104**	**1104**	**0.59**	**48.47**	**162.70**
sizeatclassdivision_mean-1sd	23.9	2802	1588	1588	1.77	16.26	54.59
sizeatclassdivision_mean+1sd	44.9	1492	846	846	0.27	107.75	361.71
sizeatclassdivision_mean-2sd	13.4	4996	2832	2832	10.02	2.87	9.63
sizeatclassdivision_mean+2sd	55.5	1209	685	685	0.14	202.36	679.34

Using the resolvable-unresolvable class boundary size value resulting from the logistic regression (34.4 pixels), the estimated resolvable range was 1.95 m (Table 3). The quotient of $L_{boundary}$ after compensating for the orientation distribution ($L_{boundary}/0.707 = 48.7$ pixels) and the median of the measured krill (42 mm) from diet samples resulted in a conversion value of 1.16 pixels/mm at the primary boundary. The resolvable range, defined by the object size at the class division boundary, was $r_{res} = 1.947$ m and the corresponding imaged volume was 0.59 m³.

The object size at the resolvable-unresolvable class boundary estimated by the logistic regression models is $L_{boundary} = 34.4$ pixels and also represents the mean of the normal distribution fit to near-boundary sizes with $sd(L_{boundary}) = 10.5$ pixels. Upper and lower 67% and 95% size intervals (mean +- 1 and 2 sd) from this fit were [23.9, 44.9] and [13.4, 55.5] pixels.

Krill density and biomass

Krill density in a given image ranged from 2 to 307 krill·m⁻³ over 26 dives (Fig 6). Mean density was 48 krill·m⁻³ (sd = 61 krill·m⁻³) and median krill density was 23.6 krill·m⁻³ (Fig 7). Estimated mean biomass density was 162.7 g·m⁻³.

Sensitivity and error

Based on the primary size-at-boundary value, $L_{boundary} = 34.4$ pixels (Table 3; Fig 8) -/+ 1*sd from the mean, 23.9 and 44.9 pixels, the estimated resolvable range and imaged volumes ranged from 2.8 m to 1.5 m and 1.77 m³ to 0.27 m³, respectively, and mean krill density estimates ranged from 16 to 108 krill·m⁻³. Based on a size-at-bound values -/+ 2*sd from the mean, 13.4 and 55.5 pixels, estimated resolvable range and imaged volumes ranged from 5.0 m to 1.2 m and 10.0 m3 to 0.14 m³, respectively, and mean krill density estimates ranged from 3 to 202 krill·m⁻³. The minimum size of resolvable objects was 14.9 pixels and resulted in a resolvable range of 4.51 m, an imaged volume of 7.4 m³, and mean mean krill density of 3.9 krill·m⁻³. The minimum size of unresolvable objects was 9.9 pixels and resulted in a resolvable range of 6.77 m and imaged volume of 25.0 m³, and mean krill density of 1.2 krill·m⁻³.

Distributions of resolvable range, imaged volume, and krill density resulting from a simulated sample near-object sizes (n = 1000, ~N(34.4, 10.5) pixels), had median values of 1.9 m, 0.58 m³, and 49.7 krill·m⁻³ (Fig 8G–8I). Mean values of each sample-estimated distribution were 2.2 m (sd = 0.8 m), 1.2 m³ (sd = 2.2 m³), 61.1 (sd = 47.9 krill·m⁻³), higher than the medians and suggesting asymmetric distributions (Fig 8D–8F). The 25th and 75th quantile values of these estimates were 1.6 and 2.4 m, 0.34 and 1.15 m³, and 25 and 83 krill·m⁻³ (Fig 8G–8I).
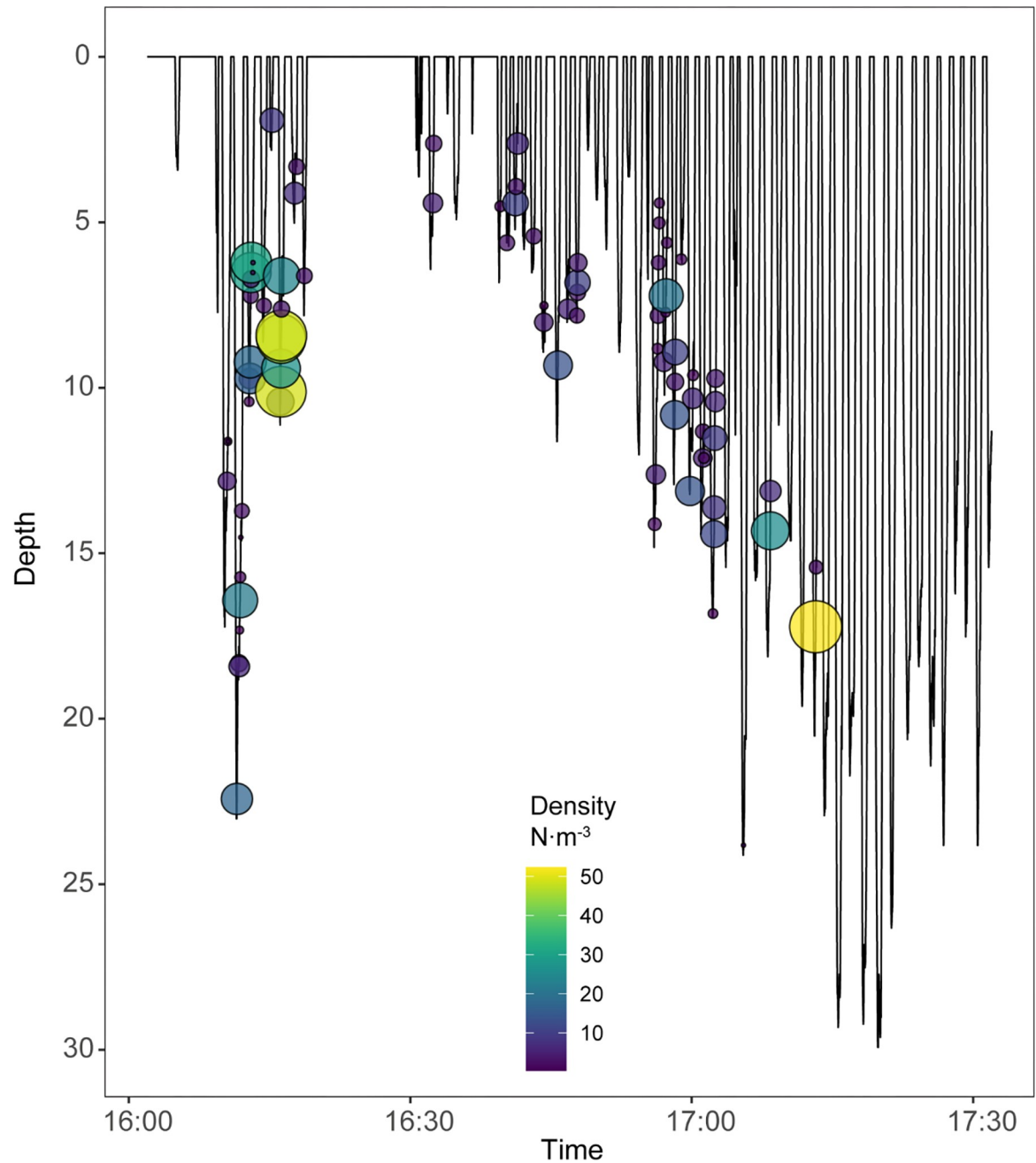
Fig 6. Antarctic krill (*Euphausia superba*) densities per image within the chinstrap penguin's (*Pygoscelis antarcticus*) dive profile for a 90-minute timespan.

https://doi.org/10.1371/journal.pone.0303633.g006

Comparison of model predictions to manual annotations

Model predictions of frame-level content match human annotations of penguin behavior such as surfacing events, underwater activity, transit periods, and prey encounters for a set of three consecutive dive events while foraging (Fig 9). For example, predictions of surface events occurred between the time of dive start and end in manual annotations while predictions for underwater events routinely occurred during dives. In between these surface events, our classifier accurately predicted imagery corresponding with the penguin head being close to the
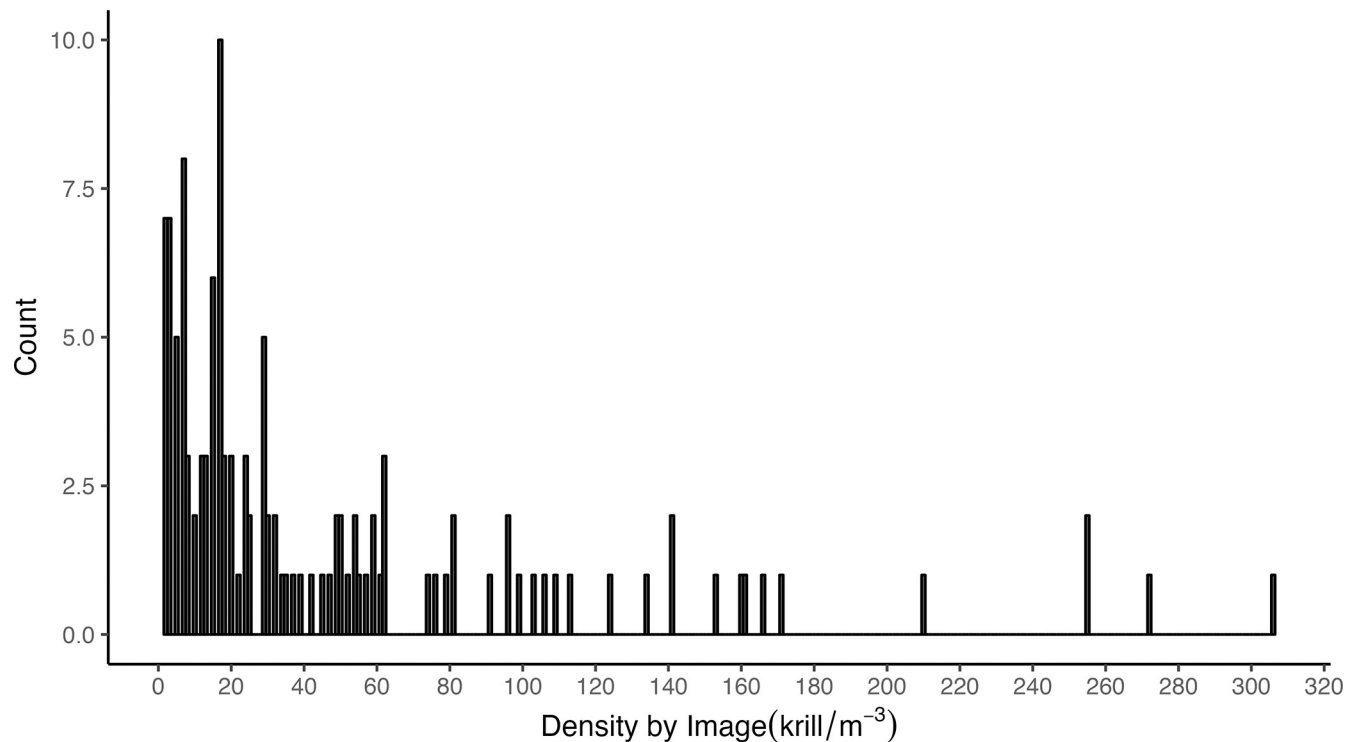
Fig 7. Histogram of estimated Antarctic krill (*Euphausia superba*) densities (n·m^{-3}) from 124 krill images predicted by the image classifier. The y-axis represents individual krill counts and the x-axis represents the density values observed from our imagery. Mean krill density was 48 krill·m^{-3} and the standard deviation was estimated as 61 krill·m^{-3}.

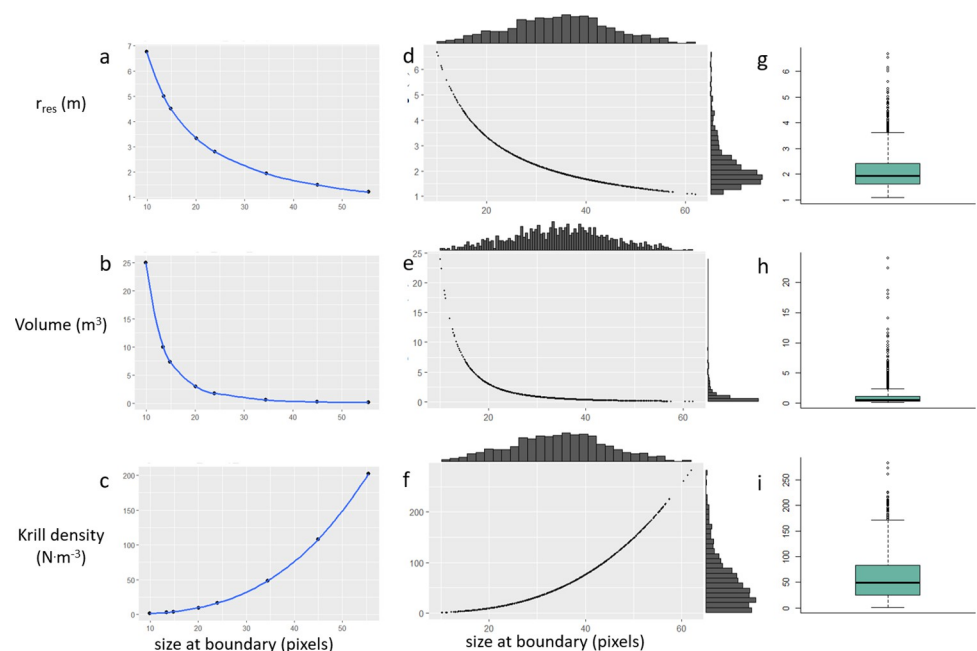**Fig 8.** (a) Resolvable range (m), (b) imaged volume (m^3), and (c), krill density (krill·m^{-3}) estimated for the discrete set of boundary values reported in Table 3. (d, e, f) Distributions and (g, h, i) boxplots of resolvable range, imaged volume, and krill density estimated by a simulated sample (n = 1000) of near-boundary object sizes.

Fig 9. Comparison of the manual annotations and VIAME predictions for a subset of dives. A) Predictions of surface (yellow circles) and underwater (blue circles) frames relative to manual observations of the start (green inverted triangle) and stop (red triangle) of individual dives. B) Predictions of krill frames (orange circles) relative to manual observations of attempted strikes at individual krill (purple triangle).

surface of the water, indicating regular breaks to breathe, or resting in place at the surface of the ocean. Our classifier also accurately predicted the transition between immediate descent or ascent (represented by the bright/dark class in our classifier) by flanking surfacing events prior to and after dives. In Fig 9, we illustrate these overlapping characteristics from our classifier model predictions and manual annotations from the video imagery.

Discussion

We used an image classifier model to automate the identification of frame-level content in animal-borne video imagery and developed a method to estimate prey (Antarctic krill) density as observed in 2D imagery. The results demonstrate that estimating volumetric prey density is possible from 2D video imagery. We note that the classifier model predicted frame-level content consistent with manual observations, highlighting the potential for automated image analysis to quantify a broad suite of penguin foraging behaviors and prey encounter events in video data. Below, we discuss our approach to automating frame-level analysis of video imagery and the several assumptions required for harnessing 2D imaging methods for 3D data collection.

Automated image analysis

We automated analysis of bio-logging video imagery collected by a chinstrap penguin foraging in the Southern Ocean. Tools such as the video loggers provide researchers with large amounts of data, which have usually been processed manually [40, 63, 64]. As methods for observing marine systems shift to automated systems [65–67] that deploy cameras, the volume of image data collected will increase to levels that exceed the ability and time of analysts to process and analyze manually. In the past, analysts relied on manual annotation [68]. Visually analyzing and manually annotating videos to identify features or events of interest is cumbersome or subject to observer bias, whereas a classifier model improves efficiency and objectivity. Resources like VIAME, that integrate annotation tools with functionality for model training, testing, and application to novel data provide convenient workflows to reduce processing bottlenecks.

Here, our model for frame-level content performed well (Table 2) despite generally low numbers of training images, highlighting the power of the EfficientNet deep learning model to detect gross differences in image content specific to our purpose. We trained the model to characterize video frames based on the environments, behaviors of penguins, and predator-prey interactions. In particular, we were interested primarily in the ability of the model to identify frames that contained krill swarms. The performance evaluation showed reasonably high precision in correctly identifying krill frames (83%). This means that 17% of incorrectly predicted krill imagery (34 out of 198 images) was likely due to a lack in clarity of images affected by motion-blur or faint swarms in the distance. Recall for krill was 63%, which is a key indicator in determining the model's sensitivity competence. Meanwhile, accuracy for the krill class was only 55%, and is determined by dividing the total number of correct predictions (i.e., TP and TN) divided by the total predictions. The largest discrepancy appeared to be from the krill and open water classes and may have two simple explanations. First, open water comprises most of the image content for these classes. Based on the results of these evaluation metrics, it is fair to assume the model could not differentiate between krill and open water at times. This is reasonable because some of the imagery we used contained large swarms of unresolvable krill in the distance, making it difficult to discern at times. Second, if krill images are affected by motion blur, or if krill targets are at large ranges or have low contrast given ambient lighting they present as small, poorly-resolved targets in the distance and difficult to identify even by human analysts. The model was sufficiently accurate for our purpose to identify the timing of krill encounters and conditions associated with basic diving behavior (Fig 9). Our results highlight the potential for future development of similar models to assess diving and foraging behaviors from video imagery.

Estimating density from 2D imagery

The process of estimating prey density from 2D animal-borne video imagery requires several assumptions about the size distribution and orientation of targets relative to the camera. With

a single lens, we cannot directly measure distance between the camera and individual krill targets. To extrapolate from imaged lengths to real-world lengths, we assumed that all krill observed are equivalent to the median length of krill found in the stomach contents of other penguins. This assumption is reasonable given the narrow size range of krill typically consumed by chinstrap penguins [61] and the consistent tracking of variation in krill sizes by different penguin species over time [69]. Additionally, variation in the orientation of krill relative to the camera can affect the estimates of target lengths in the image. For example, if all krill were observed axially (i.e., only head or tail face the camera), then estimating body length would be impractical. Based on the subset of imagery we used, these extreme orientations were rare. However, most individuals were observed in variable orientations, likely due to escape behaviors and scattering in multiple directions as a predator approached. It was difficult to discern if krill were directly perpendicular to the camera as well, therefore, the variation in orientations necessitated a mechanism to adjust imaged lengths for potential angular shifts. We therefore assumed that krill orientation angles were distributed randomly with a uniform distribution. These assumptions provide the basis to estimate the distance of targets from the lens and construction of 3D information from the video image.

Environmental factors may also affect the estimate of imaged volume, including water column conditions (e.g., visibility, lighting, or turbidity). Water column conditions like visibility, lighting, or turbidity can vary, particularly when imagery is captured at different depths or at different times of the day when solar incidence differs. For example, if krill foraging occurred at greater depths or was recorded near sunset, it is likely that our ability to resolve targets would weaken as light levels fall, effectively reducing the volume of the image. Here, our methodology is limited in that we worked to minimize such effects by focusing on well-lit surface waters <30m in depth (Fig 6) within a short span of foraging activities during midday. Water clarity within these was also similar across the short time span for which video data were available. Our estimate of imaged volume may not be applicable across a broad spectrum of conditions due to the relatively small size of the dataset. This is not a large enough sample size of imagery containing observations from near to far; nor is there enough variation in water column conditions. This method can, nonetheless, address density estimates through assessment of the resolvability of targets in such images.

Another key assumption in our method is how to identify the appropriate cut-off for the size of resolvable krill that represent the limit of the resolvable range. The concept of resolvable range considers that resolvability (or focus) of targets is limited primarily by distance from the camera. In an ideal case without image blur or distortion, resolvable targets would be located near the camera and appear larger while unresolvable targets would be farther away and smaller, assuming that the target size distribution did not vary with range. The resolvable range could then be estimated based on the upper size limit of the unresolvable targets or the lower size limit of resolvable targets. However, in our data there was considerable overlap of the smallest resolvable and unresolvable targets (Fig 5). We assume that many targets are unresolvable for reasons other than distance from the camera, including motion blur. Therefore, we focused our method on the estimated mean from a binary classifier based on a logistic regression fit to sizes of all unresolvable and resolvable objects at defined boundaries between these classes. This normal model fit to near boundary size distributions also provided a probability density function that could be applied to any estimated value of resolvable range, volume, and density.

We estimated density using an imaged volume estimated by resolvable range (essentially depth of field). We use the "resolvability" class data in conjunction with size distribution of objects to estimate the range. While we have an estimate of the range that defines the transition from resolvable to unresolvable objects, we acknowledge the potential for errors of assignment

(subjectivity) by one or more analysts. The data and results of the process of assigning targets to a binary scheme demonstrate that there is no clear and strict planar boundary that defines the far edge of the imaged volume. Targets do not transition from being resolvable to unresolvable suddenly over a small change of range. The data reinforce that there is a marginal region, not a plane, where visual characteristics change from sharp and detailed to unidentifiable. We provide estimates of resolvable range, imaged volume, and density for the full range of target sizes that we consider relevant to the process as supported by the data. Our results suggest that the resolvable range is 2m, imaged volume is 0.6 m^3, and krill density is 50 krill·m^{-3}, approximately. The 25th and 75th quantile values of these estimates were within a factor of two times the median values, at 1.6 and 2.4 m, 0.34 and 1.15 m^3, and 25 and 83 krill·m^{-3}, respectively (Fig 8G–8I). Within this transitional region, maybe we could classify targets as marginally resolvable and use their size distribution to estimate the range. Indeed, our initial efforts at target classification included a marginal class, but we determined that it was not feasible to reliably identify objects as marginal. Therefore, we chose the binary scheme, which ultimately allowed us to estimate the resolvable range and imaged volume using the logistic regression classifier, and to characterize the extent of the nebulous boundary.

Similarly, the size distribution of the unresolvable class alone may be enough to estimate the boundary. The sizes of unresolvable objects had a mode of 25, median of 40, mean of 45, and standard deviation of 25 pixels (Fig 5). Given the asymmetry, perhaps the mode or median could serve as alternative estimates of near-boundary object size. The mode of unresolvable object sizes is 9 pixels smaller than the size at boundary estimated from logistic regression. As such, the estimates for range, volume, and density are, respectively, 1.4, 3.4, and 0.4 times the values produced using $L_{boundary}$ = 34.4 pixels (Fig 8 and Table 3). Similarly, the median unresolvable object size is 5 pixels larger and relates to estimates that are 0.9, 1.9, and 1.5 times the same. Unresolvable objects had a wider distribution of sizes than what we used based on assuming that all objects with sizes smaller than the class boundary size from logistic regression were technically unresolvable (and mirroring those to characterize the full marginal region). Therefore, using only the distribution of unresolvable targets is not recommended, not only for this reason, but also because many unresolvable targets are strongly blurred krill at close range within the imaged volume. This also suggests that we adhered to assignment of objects to resolvable-unresolvable classes based on visual characteristics of targets and not context, which is more difficult than it may seem. The size distribution of the resolvable targets overlapped with that of the unresolvable objects, suggesting that some of the resolvable targets were probably truly unresolvable, and also perhaps some smaller detail or context led to similar misassignment in the other direction.

We allowed the classified object size data to determine the boundary despite the likelihood that the data probably contained some errors (e.g. large unresolvable targets that existed in the volume). We did not arbitrarily filter the data because we believed that some errors like this existed. Therefore, the boundary size value based on logistic regression is larger and the estimated resolvable range is smaller than a boundary defined on the smallest resolvable targets, which would ideally be located at the limit of the imaged volume. The steep curves in Fig 5 demonstrate the sensitivity of resolvable range and imaged volume estimates to the changes of the smallest, presumably distant, object sizes. For example, the minimum unresolvable object size (9.9 pixels in length) represents the smallest visually distinguishable objects (presumed to be krill and not bubbles) at the largest range that would be considered to represent an imaged volume encompassing all objects. The resolvable range estimated based on the smallest unresolvable objects was 6.7 m and imaged volume was 25.0 m^3. The minimum resolvable object size (14.9 pixels in length) represents the smallest identifiable krill. The values of resolvable range and volume associated with this size were 4.5 m and 7.4 m. However, several key points

support it is not justified to define a boundary based on the extremes from either class: the probability distribution of near-boundary and unresolvable object sizes; the difficulty for an analyst to distinguish and mark such targets for measurement; and that such small and indistinct objects are not identifiable.

Someone attempting to apply this method to their own imagery should not discount the difficulty in assigning targets to the resolvable and unresolvable classes. There is subjectivity involved, and target details and overall appearance can be strongly affected by factors such as motion blur, irregular illumination, or lighting deficiencies that require the analyst to assign a target to the unresolvable class if they are adhering strictly to the definition. Quantitative methods such as intensity gradients of targets may be applied to classify objects as resolvable/unresolvable, or in focus/not in focus [70], but those gradients would be diminished for objects that are affected by strong motion blur, for instance, leading to the same conclusion as the human visual analyst. Expert annotations are critical to the process, here and in general, for generating datasets that serve as a basis for training more complicated models. Augmented annotations based on intensity gradient metrics for example, would also require careful considerations and perhaps filtering if a trained model is to be trusted. Platforms with moving targets such as this dataset remain a challenge.

We further tested the validity of our biomass estimate relative to verified acoustic survey biomass estimates from the same time period [67]. Currently, there are a lack of studies available to affirm relative thickness of krill swarms actively being foraged by penguins, however, there are some that have estimated swarm vertical thickness [71, 72]. Swarm types 1 and 2 of [72] seemed to represent the swarm types similar to what have been observed by acoustic surveys in our study area [67]. Mean thickness of swarms according to [71] and the type 2 swarms of [72] was 10 m, and mean thickness of the type 1 swarms of [72] was 5 m. Assuming that the penguins in our study dove past the bottom of the swarms before ascending, mean thickness according to the penguins was approximately 7 m (Fig 9). These various krill swarm thicknesses were considered in order to estimate volumetric biomass densities from areal biomass estimates.

The average areal biomass density during January in the area (based on glider acoustic survey [67]) was 52.9 g·m^{-2}. Assuming a swarm thickness of either 5 m [71] or 10 m [72], the areal biomass estimate from [67] converts to volumetric biomass densities of 10.57 g·m^{-3} and 5.3 g·m^{-3}, respectively. The volumetric biomass density estimated from our imagery (krill counts and estimated imaged volume) was 26.1 g·m^{-3}, which is about 2.5 times larger than acoustic survey estimates from the previous year [67] assuming a 5-m thick swarm, and about 5 times the density assuming 10-m thick swarms. If we use the average foraging depths from our study (Fig 9), and assume that swarms were 7 m thick, then the volumetric biomass estimate of [67] was 7.6 g·m^{-3}, or 3.4 times less than our image based estimate. Biomass density values from repeated acoustic surveys one month apart in this area differed by a factor of 2 [67]. Given the limitations in vertical swarm thickness, and large acoustic estimate discrepancies between the austral months, our mean volumetric biomass is valid. This confirms the ability in which this methodology can estimate local krill density at a scale that has not yet been observed. Yet, it is important to emphasize our limited study is not representative of general krill population patterns and future efforts should consider utilizing a larger dataset over a larger timescale to discuss broader implications.

Conclusion

We estimated krill density from an arbitrarily-chosen subset of single camera images collected from the perspective of one penguin. Consequently, we stress that the data reported here are

not representative of the krill population as a whole, nor representative of the full range in size and density of krill swarms likely to be encountered during a foraging trip. Rather, we highlight that the method is capable of assessing a wide range of swarm densities that a penguin may encounter. The results provide insight to foraging patterns of individual penguins and their prey availability. For example, krill density encountered by a foraging penguin can change dramatically within and between dives (Fig 6). Such small-scale variation in prey density on short time and space scales suggests a high degree of patchiness in krill swarms that likely impacts foraging success and effort by air breathing predators like penguins.

The limitations of single image and absence of depth data has required us to estimate range and volume based on target object characteristics with possibly much more uncertainty than if we could measure depth directly. Perhaps there are other options for cameras and lenses that could achieve a sharper boundary division, a more discrete depth of field, and more importantly, direct measurement of distance (range) to objects. Range and target size estimates could be accomplished by using a depth sensor next to the camera (as in RGB-D) or stereo camera systems. A system that could create an accurate, dense 3D reconstruction of the scene at high speed and simultaneously track each target to enable identification and measurement would be valuable. Such a system would facilitate the ability of more detailed quantification of predator-prey interactions. However, we are not aware of any available systems that are small enough, with a small enough mass, to be borne by small marine organisms. Motion blur, caused by simultaneous high-speed motion and rapid changes of direction by the predator and prey, is a big problem which might be addressed by using a high intensity short duration strobe, but in shallow, clear water during daytime, the efficacy of such a strobe may be diminished. Therefore, very high-speed image acquisition rates and very short exposure durations may be preferable. We would encourage the development of such hardware that could be borne by small marine animals like penguins. Until then, the method we describe may be used to estimate imaged volume and prey density from single cameras.

Future work to examine the links between estimates of prey density with consumption rates, foraging behavior (e.g., decisions to continue foraging locally or transit to new foraging areas), and the energetic expenses associated with those decisions are likely to yield important insight in how real-time variation in prey density affects predators. Additionally, prey density data from predator-borne systems may provide information important to managing krill fisheries, such as how well the density estimated from surveys using ships or autonomous vehicles [70] represents the prey directly available to predators. Furthermore, with a changing climate, data such as this may provide a first-hand predator perspective on changes in prey availability. The methods used here to estimate volumetric density from 2D imagery should be generally applicable to other open-water systems with swarming prey.

Acknowledgments

Author Contributions

Conceptualization: Victoria R. Hermanson, George R. Cutter, Jefferson T. Hinke, George M. Watters.

Data curation: Victoria R. Hermanson.

Formal analysis: Victoria R. Hermanson, George R. Cutter, Jefferson T. Hinke.

Investigation: Victoria R. Hermanson, George R. Cutter.

Methodology: Victoria R. Hermanson, George R. Cutter, Jefferson T. Hinke.

Project administration: Victoria R. Hermanson.

Resources: Jefferson T. Hinke, George M. Watters.

Software: Matthew Dawkins.

Supervision: George R. Cutter, Jefferson T. Hinke, George M. Watters.

Visualization: Victoria R. Hermanson, George R. Cutter.

Writing – original draft: Victoria R. Hermanson.

Writing – review & editing: George R. Cutter, Jefferson T. Hinke.

References

1. Williams CL, Ponganis PJ. Diving physiology of marine mammals and birds: the development of biologging techniques. Philosophical Transactions of the Royal Society B. 2021 Aug 2; 376(1830):20200211. https://doi.org/10.1098/rstb.2020.0211 PMID: 34121464

2. Wilmers CC, Nickel B, Bryce CM, Smith JA, Wheat RE, Yovovich V. The golden age of bio-logging: How animal-borne sensors are advancing the frontiers of ecology. Ecology. 2015 Jul; 96(7):1741–53. https://doi.org/10.1890/14-1401.1 PMID: 26378296

3. Kooyman GL, Cherel Y, Maho YL, Croxall JP, Thorson PH, Ridoux V, et al. Diving behavior and energetics during foraging cycles in king penguins. Ecological Monographs. 1992 Mar; 62(1):143–63.

4. Chappell MA, Shoemaker VH, Janes DN, Maloney SK, Bucher TL. Energetics of foraging in breeding Adélie penguins. Ecology. 1993 Dec; 74(8):2450–61.

5. Shaffer SA, Tremblay Y, Weimerskirch H, Scott D, Thompson DR, Sagar PM, et al. Migratory shearwaters integrate oceanic resources across the Pacific Ocean in an endless summer. Proceedings of the National Academy of Sciences. 2006 Aug 22; 103(34):12799–802.

6. Block BA, Jonsen ID, Jorgensen SJ, Winship AJ, Shaffer SA, Bograd SJ, et al. Tracking apex marine predator movements in a dynamic ocean. Nature. 2011 Jul 7; 475(7354):86–90. https://doi.org/10.1038/nature10082 PMID: 21697831

7. Grémillet D, Kuntz G, Delbart F, Mellet M, Kato A, Robin JP, et al. Linking the foraging performance of a marine predator to local prey abundance. Functional Ecology. 2004 Dec 1:793–801.

8. Pickett EP, Fraser WR, Patterson-Fraser DL, Cimino MA, Torres LG, Friedlaender AS. Spatial niche partitioning may promote coexistence of Pygoscelis penguins as climate-induced sympatry occurs. Ecology and Evolution. 2018 Oct; 8(19):9764–78. https://doi.org/10.1002/ece3.4445 PMID: 30386573

9. Raymond B, Lea MA, Patterson T, Andrews-Goff V, Sharples R, Charrassin JB, et al. Important marine habitat off east Antarctica revealed by two decades of multi-species predator tracking. Ecography. 2015 Feb; 38(2):121–9.

10. Hindell MA, Reisinger RR, Ropert-Coudert Y, Hückstädt LA, Trathan PN, Bornemann H, et al. Tracking of marine predators to protect Southern Ocean ecosystems. Nature. 2020 Apr 2; 580(7801):87–92. https://doi.org/10.1038/s41586-020-2126-y PMID: 32238927

11. Naito Y, Bornemann H, Takahashi A, McIntyre T, Plötz J. Fine-scale feeding behavior of Weddell seals revealed by a mandible accelerometer. Polar Science. 2010 Aug 1; 4(2):309–16.

12. Kokubun N, Kim JH, Shin HC, Naito Y, Takahashi A. Penguin head movement detected using small accelerometers: a proxy of prey encounter rate. Journal of Experimental Biology. 2011 Nov 15; 214 (22):3760–7. https://doi.org/10.1242/jeb.058263 PMID: 22031740

13. Viviant M, Jeanniard-du-Dot T, Monestiez P, Authier M, Guinet C. Bottom time does not always predict prey encounter rate in Antarctic fur seals. Functional Ecology. 2016 Nov; 30(11):1834–44.

14. Horsburgh J, Morrice M, Lea MA, Hindell M. Determining feeding events and prey encounter rates in a southern elephant seal: a method using swim speed and stomach temperature. Marine Mammal Science. 2008 Jan 1; 24(1):207–17.

15. Weimerskirch H, Robertson G. Satellite tracking of light-mantled sooty albatrosses. Polar Biology. 1994 Feb; 14:123–6.

16. Robinson PW, Tremblay Y, Crocker DE, Kappes MA, Kuhn CE, Shaffer SA, et al. A comparison of indirect measures of feeding behaviour based on ARGOS tracking data. Deep Sea Research Part II: Topical Studies in Oceanography. 2007 Feb 1; 54(3–4):356–68.

17. Parrish FA, Craig MP, Ragen TJ, Marshall GJ, Buhleier BM. Identifying diurnal foraging habitat of endangered Hawaiian monk seals using a seal-mounted video camera. Marine Mammal Science. 2000 Apr; 16(2):392–412.

18. Parrish FA, Abernathy K, Marshall GJ, Buhleier BM. Hawaiian monk seals (Monachus schauinslandi) foraging in deep-water coral beds. Marine Mammal Science. 2002 Jan; 18(1):244–58.

19. Littnan CL, Baker JD, Parrish FA, Marshall GJ. Effects of video camera attachment on the foraging behavior of immature Hawaiian monk seals. Marine Mammal Science. 2004 Apr; 20(2):345–52.

20. Parrish FA, Marshall GJ, Littnan C, Heithaus M, Canja S, Becker B, et al. Foraging of juvenile monk seals at French Frigate Shoals, Hawaii. Marine Mammal Science. 2005 Jan; 21(1):93–107.

21. Friedlaender AS, Lawson GL, Halpin PN. Evidence of resource partitioning between humpback and minke whales around the western Antarctic Peninsula. Marine mammal science. 2009 Apr; 25(2):402–15.

22. Santora JA, Reiss CS, Loeb VJ, Veit RR. Spatial association between hotspots of baleen whales and demographic patterns of Antarctic krill Euphausia superba suggests size-dependent predation. Marine Ecology Progress Series. 2010 Apr 29; 405:255–69.

23. Potvin J, Cade DE, Werth AJ, Shadwick RE, Goldbogen JA. A perfectly inelastic collision: bulk prey engulfment by baleen whales and dynamical implications for the world's largest cetaceans. American Journal of Physics. 2020 Oct 1; 88(10):851–63.

24. Friedlaender AS, Joyce T, Johnston DW, Read AJ, Nowacek DP, Goldbogen JA, et al. Sympatry and resource partitioning between the largest krill consumers around the Antarctic Peninsula. Marine Ecology Progress Series. 2021 Jul 8; 669:1–6.

25. Nichols RC, Cade DE, Kahane-Rapport S, Goldbogen J, Stimpert A, Nowacek D, et al. Intra-seasonal variation in feeding rates and diel foraging behaviour in a seasonally fasting mammal, the humpback whale. Royal Society Open Science. 2022 Jul 6; 9(7):211674. https://doi.org/10.1098/rsos.211674 PMID: 35814912

26. Thiebot JB, Arnould JP, Gómez-Laich A, Ito K, Kato A, Mattern T, et al. Jellyfish and other gelata as food for four penguin species–insights from predator-borne videos. Frontiers in Ecology and the Environment. 2017 Oct; 15(8):437–41.

27. Mattern T, McPherson MD, Ellenberg U, van Heezik Y, Seddon PJ. High definition video loggers provide new insights into behaviour, physiology, and the oceanic habitat of a marine predator, the yellow-eyed penguin. PeerJ. 2018 Sep 19; 6:e5459. https://doi.org/10.7717/peerj.5459 PMID: 30258706

28. Sabrina H, Gabriela S, Klemens P, Thomas M, Andrea RR. Niche partitioning between coexisting gentoo Pygoscelis papua and Magellanic penguins Spheniscus magellanicus at Martillo Island, Argentina. Marine Biology. 2020 Aug; 167:1–0.

29. Pistorius PA, Green DB, Seddon PJ, Thiebault A. In situ observation of a record-sized squid prey consumed by a gentoo penguin. Polar Biology. 2020 Mar; 43:279–83.

30. Harris S, Pütz K, Mattern T, Scioscia G, Raya Rey A. The role of conspecifics during pelagic foraging of Magellanic and benthic foraging of Gentoo penguins in the Beagle Channel, Argentina. Marine Biology. 2023 Feb; 170(2):17.

31. Carroll G, Slip D, Jonsen I, Harcourt R. Supervised accelerometry analysis can identify prey capture by penguins at sea. Journal of Experimental Biology. 2014 Dec 15; 217(24):4295–302. https://doi.org/10.1242/jeb.113076 PMID: 25394635

32. Handley JM, Thiebault A, Stanworth A, Schutt D, Pistorius P. Behaviourally mediated predation avoidance in penguin prey: in situ evidence from animal-borne camera loggers. Royal Society open science. 2018 Aug 22; 5(8):171449. https://doi.org/10.1098/rsos.171449 PMID: 30224976

33. Davis RW, Fuiman LA, Williams TM, Collier SO, Hagey WP, Kanatous SB, et al. Hunting behavior of a marine mammal beneath the Antarctic fast ice. Science. 1999 Feb 12; 283(5404):993–6. https://doi.org/10.1126/science.283.5404.993 PMID: 9974394

34. Krause DJ, Goebel ME, Marshall GJ, Abernathy K. Novel foraging strategies observed in a growing leopard seal (Hydrurga leptonyx) population at Livingston Island, Antarctic Peninsula. Animal Biotelemetry. 2015 Dec; 3(1):1–4.

35. Akiyama Y, Akamatsu T, Rasmussen MH, Iversen MR, Iwata T, Goto Y, et al. Leave or stay? Video-logger revealed foraging efficiency of humpback whales under temporal change in prey density. PloS one. 2019 Feb 5; 14(2):e0211138. https://doi.org/10.1371/journal.pone.0211138 PMID: 30721236

36. Ponganis PJ, Dam RV, Marshall G, Knower T, Levenson DH. Sub-ice foraging behavior of emperor penguins. Journal of Experimental Biology. 2000 Nov 1; 203(21):3275–8. https://doi.org/10.1242/jeb.203.21.3275 PMID: 11023847

37. Takahashi A, Sato K, Naito Y, Dunn MJ, Trathan PN, Croxall JP. Penguin–mounted cameras glimpse underwater group behaviour. Proceedings of the Royal Society of London. Series B: Biological Sciences. 2004 Aug 7; 271(suppl_5):S281–2.

38. Zhang K, Yang M, Lang SD, McInnes AM, Sherley RB, Burghardt T. Diving with Penguins: Detecting Penguins and their Prey in Animal-borne Underwater Videos via Deep Learning. arXiv preprint arXiv:2308.07267. 2023 Aug 14.

39. Watanabe YY, Takahashi A. Linking animal-borne video to accelerometers reveals prey capture variability. Proceedings of the National Academy of Sciences. 2013 Feb 5; 110(6):2199–204. https://doi.org/10.1073/pnas.1216244110 PMID: 23341596

40. Hinke JT, Russell TM, Hermanson VR, Brazier L, Walden SL. Serendipitous observations from animal-borne video loggers reveal synchronous diving and equivalent simultaneous prey capture rates in chinstrap penguins. Marine Biology. 2021 Aug; 168:1–2.

41. Manco F, Lang SD, Trathan PN. Predicting foraging dive outcomes in chinstrap penguins using biologging and animal-borne cameras. Behavioral Ecology. 2022 Sep 1; 33(5):989–98.

42. Weinstein BG. A computer vision for animal ecology. Journal of Animal Ecology. 2018 May; 87(3):533–45. https://doi.org/10.1111/1365-2656.12780 PMID: 29111567

43. Zhang S, Yang X, Wang Y, Zhao Z, Liu J, Liu Y, et al. Automatic fish population counting by machine vision and a hybrid deep neural network model. Animals. 2020 Feb 24; 10(2):364. https://doi.org/10.3390/ani10020364 PMID: 32102380

44. Mensink T, Verbeek J, Perronnin F, Csurka G. Metric learning for large scale image classification: Generalizing to new classes at near-zero cost. InComputer Vision–ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7–13, 2012, Proceedings, Part II 12 2012 (pp. 488–501). Springer Berlin Heidelberg.

45. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. Imagenet large scale visual recognition challenge. International journal of computer vision. 2015 Dec; 115:211–52.

46. Dawkins M, Sherrill L, Fieldhouse K, Hoogs A, Richards B, Zhang D, et al. 2017. An open-source platform for underwater image and video analytics 2017 IEEE Winter Conference on Applications of Computer Vision. pp. (898–906).

47. Chiang JY, Chen YC, Chen YF. Underwater image enhancement: using wavelength compensation and image dehazing (WCID). InAdvanced Concepts for Intelligent Vision Systems: 13th International Conference, ACIVS 2011, Ghent, Belgium, August 22–25, 2011. Proceedings 13 2011 (pp. 372–383). Springer Berlin Heidelberg.

48. Cao Y, Fang S, Wang F. Single image multi-focusing based on local blur estimation. In2011 Sixth International Conference on Image and Graphics 2011 Aug 12 (pp. 168–175). IEEE.

49. Tang H, Cohen S, Price B, Schiller S, Kutulakos KN. Depth from defocus in the wild. In Proceedings of the IEEE conference on computer vision and pattern recognition 2017 (pp. 2740–2748).

50. Bailey SW, Echevarria JI, Bodenheimer B, Gutierrez D. Fast depth from defocus from focal stacks. The Visual Computer. 2015 Dec; 31:1697–708.

51. Carvalho M, Le Saux B, Trouvé-Peloux P, Almansa A, Champagnat F. Deep depth from defocus: how can defocus blur improve 3D estimation using dense neural networks?. In Proceedings of the European Conference on Computer Vision (ECCV) Workshops 2018 (pp. 0–0).

52. Zelenka C, Koch R. Single Image Plankton 3D Reconstruction from Extended Depth of Field Shadowgraph. In Pattern Recognition and Information Forensics: ICPR 2018 International Workshops, CVAUI, IWCF, and MIPPSNA, Beijing, China, August 20–24, 2018, Revised Selected Papers 24 2019 (pp. 76–85). Springer International Publishing.

53. Gur S, Wolf L. Single image depth estimation trained via depth from defocus cues. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2019 (pp. 7683–7692).

54. Zhuo S, Sim T. Defocus map estimation from a single image. Pattern Recognition. 2011 Sep 1; 44(9):1852–8.

55. Mertan A, Duff DJ, Unal G. Single image depth estimation: An overview. Digital Signal Processing. 2022 Apr 30; 123:103441.

56. Tan M, Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In International conference on machine learning 2019 May 24 (pp. 6105–6114). PMLR.

57. Kuhn M. The Caret package. 17 Measuring Performance. 2019.

58. Goodwin M, Halvorsen KT, Jiao L, Knausgård KM, Martin AH, Moyano M, et al. Unlocking the potential of deep learning for marine ecology: overview, applications, and outlook. ICES Journal of Marine Science. 2022 Mar; 79(2):319–36.

59. Hinke JT, Giuseffi LM, Hermanson VR, Woodman SM, Krause DJ. Evaluating Thermal and Color Sensors for Automating Detection of Penguins and Pinnipeds in Images Collected with an Unoccupied Aerial System. Drones. 2022 Sep 15; 6(9):255.

60. Wilson RP, Pütz K, Grémillet D, Culik BM, Kierspel M, Regel J, et al. Reliability of stomach temperature changes in determining feeding characteristics of seabirds. Journal of Experimental Biology. 1995 May 1; 198(5):1115–35. https://doi.org/10.1242/jeb.198.5.1115 PMID: 9318941

61. Miller AK, Trivelpiece WZ. Cycles of Euphausia superba recruitment evident in the diet of Pygoscelid penguins and net trawls in the South Shetland Islands, Antarctica. Polar Biology. 2007 Nov; 30:1615–23.

62. Siegel V, Reiss CS, Dietrich KS, Haraldsson M, Rohardt G. Distribution and abundance of Antarctic krill (Euphausia superba) along the Antarctic Peninsula. Deep sea research part I: oceanographic research papers. 2013 Jul 1; 77:63–74.

63. Malde K, Handegard NO, Eikvil L, Salberg AB. Machine intelligence and the data-driven future of marine science. ICES Journal of Marine Science. 2020 Jul; 77(4):1274–85.

64. Underwood MJ, Rosen S, Engås A, Eriksen E. Deep vision: An in-trawl stereo camera makes a step forward in monitoring the pelagic community. PLos One. 2014 Nov 13; 9(11):e112304. https://doi.org/10.1371/journal.pone.0112304 PMID: 25393121

65. Greene CH, Meyer-Gutbrod EL, McGarry LP, Hufnagle LC Jr, Chu D, McClatchie S, et al. A wave glider approach to fisheries acoustics: transforming how we monitor the nation's commercial fisheries in the 21st century. Oceanography. 2014 Dec 1; 27(4):168–74.

66. Meinig C, Lawrence-Slavas N, Jenkins R, Tabisola HM. The use of Saildrones to examine spring conditions in the Bering Sea: Vehicle specification and mission performance. InOCEANS 2015-MTS/IEEE Washington 2015 Oct 19 (pp. 1–6). IEEE.

67. Reiss CS, Cossio AM, Walsh J, Cutter GR, Watters GM. Glider-Based estimates of meso-zooplankton biomass density: a fisheries case study on antarctic krill (Euphausia superba) around the northern antarctic peninsula. Frontiers in Marine Science. 2021 Mar 25; 8:604043.

68. Schneider S, Greenberg S, Taylor GW, Kremer SC. Three critical factors affecting automated image species recognition performance for camera traps. Ecology and evolution. 2020 Apr; 10(7):3503–17. https://doi.org/10.1002/ece3.6147 PMID: 32274005

69. Hinke JT, Salwicka K, Trivelpiece SG, Watters GM, Trivelpiece WZ. Divergent responses of Pygoscelis penguins reveal a common environmental driver. Oecologia. 2007 Oct; 153:845–55. https://doi.org/10.1007/s00442-007-0781-4 PMID: 17566778

70. Reiss CS, Cossio AM, Loeb V, Demer DA. Variations in the biomass of Antarctic krill (Euphausia superba) around the South Shetland Islands, 1996–2006. ICES Journal of Marine Science. 2008 May 1; 65(4):497–508.

71. Ryabov AB, Tarling GA. Scaling of size, shape and surface roughness in Antarctic krill swarms. ICES Journal of Marine Science. 2019 Jul 1; 76(4):1177–88.

72. Tarling GA, Klevjer T, Fielding S, Watkins J, Atkinson A, Murphy E, et al. Variability and predictability of Antarctic krill swarm structure. Deep Sea Research Part I: Oceanographic Research Papers. 2009 Nov 1; 56(11):1994–2012.