

## RESEARCH ARTICLE

## Aerial-trained deep learning networks for surveying cetaceans from satellite imagery

Alex Borowicz<sup>1,2\*</sup>, Hieu Le<sup>2,3</sup>, Grant Humphries<sup>4</sup>, Georg Nehls<sup>5</sup>, Caroline Höschle<sup>5</sup>, Vladislav Kosarev<sup>5</sup>, Heather J. Lynch<sup>1,2</sup>

**1** Department of Ecology & Evolution, Stony Brook University, Stony Brook, New York, United States of America, **2** Institute for Advanced Computational Science, Stony Brook University, Stony Brook, New York, United States of America, **3** Department of Computer Science, Stony Brook University, Stony Brook, New York, United States of America, **4** HiDef Aerial Surveying Ltd., Cleator Moor, Cumbria, United Kingdom, **5** BioConsult SH GmbH & Co. KG, Husum, Germany

\* [aborowicz@coa.edu](mailto:aborowicz@coa.edu)

## Abstract

Most cetacean species are wide-ranging and highly mobile, creating significant challenges for researchers by limiting the scope of data that can be collected and leaving large areas un-surveyed. Aerial surveys have proven an effective way to locate and study cetacean movements but are costly and limited in spatial extent. Here we present a semi-automated pipeline for whale detection from very high-resolution (sub-meter) satellite imagery that makes use of a convolutional neural network (CNN). We trained ResNet, and DenseNet CNNs using down-scaled aerial imagery and tested each model on 31 cm-resolution imagery obtained from the WorldView-3 sensor. Satellite imagery was tiled and the trained algorithms were used to classify whether or not a tile was likely to contain a whale. Our best model correctly classified 100% of tiles with whales, and 94% of tiles containing only water. All model architectures performed well, with learning rate controlling performance more than architecture. While the resolution of commercially-available satellite imagery continues to make whale identification a challenging problem, our approach provides the means to efficiently eliminate areas without whales and, in doing so, greatly accelerates ocean surveys for large cetaceans.

## OPEN ACCESS

**Citation:** Borowicz A, Le H, Humphries G, Nehls G, Höschle C, Kosarev V, et al. (2019) Aerial-trained deep learning networks for surveying cetaceans from satellite imagery. *PLoS ONE* 14(10): e0212532. <https://doi.org/10.1371/journal.pone.0212532>

**Editor:** Paweł Pławiak, Politechnika Krakowska im Tadeusza Kosciuszki, POLAND

**Received:** January 31, 2019

**Accepted:** September 18, 2019

**Published:** October 1, 2019

**Copyright:** © 2019 Borowicz et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the manuscript and Supporting Information files. Satellite imagery is licensed from Digital Globe and cannot be distributed by the authors, but can be acquired directly from Digital Globe (<https://www.digitalglobe.com/>). All downsampled aerial imagery and scripts are available on GitHub (DOI: [10.5281/zenodo.3356970](https://doi.org/10.5281/zenodo.3356970)). [S9 File](#), [S10 File](#), and [S11 File](#) include the data required to reproduce plots and the statistics cited.

## Introduction

There is tremendous interest in understanding if and how cetacean populations are recovering following the cessation of intense commercial whaling, yet their extensive ranges and high level of mobility combine with the challenges of research at sea to leave them poorly studied in many regions. While many cetacean species favor shelf edges and other zones of deep-water upwelling across ocean basins [1–4], most cetacean research is focused on coastal areas where populations may be concentrated at key times of the year and are logistically easier to survey. Far less work has been done to understand patterns of cetacean habitat use along distant continental shelf regions, and limited cetacean surveys in deep-water habitat may skew our understanding about preferred habitat [5]. While data loggers and transponders have been

**Funding:** This work was supported by the Directorate for Computer and Information Science and Engineering grant 1531492 and by a satellite imagery grant from the Digital Globe Foundation (<http://foundation.digitalglobe.com/>) to A.B. and through a kick-starter grant from the European Space Agency (<http://www.esa.int/ESA>) to G.N. BioConsult SH GmbH & Co. (<https://bioconsult-sh.de/en/>). G.N., C.H., and V.K are employed by BioConsult and G.H. is employed by HiDef Aerial Surveying Ltd. (<http://www.hidefsurveying.com/>), which provided aerial imagery and is a subsidiary of BioConsult. A sub-award from BioConsult to the Stony Brook Research Foundation supported H. Le and H. Lynch's contributions to this analysis. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** I have read the journal's policy and the authors of this manuscript have the following competing interests: GH, GN, CH, and VK work for companies which may consider a commercial application for this research in the future. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

employed for many species to track movements far outside the range of direct observation and provide data with high spatial and temporal resolution [6–10], the cost of these devices often drastically limits the number of animals that might be tracked.

Broad-scale or basin-scale surveys are exceedingly challenging and costly, regardless of modality. Between 1975 and 2005, only 25% of the world's oceans were surveyed for cetaceans, with a high proportion of surveys falling inside the territorial waters of the United States [5,11]. Basin-scale surveys are unavoidably multi-year efforts in which a region is surveyed in parts over consecutive years or involve extensive mark-recapture studies [12–14]. Another approach has been to combine data from heterogeneous sources such as aerial and ship surveys (e.g., [15]), strandings, and whaling data [16]. However, these methods are expensive and the results are difficult to interpret due to large data gaps. New methods that complement existing tools and address these challenges of scale are required. While remote sensing has been used for decades to track coarse-grained changes in the environment (e.g., sea ice, land cover, urban development), the use of satellite imagery to directly survey animals is much more recent and hinges on the use of very high-resolution (sub-meter) imagery that can capture individual animals on the landscape (e.g., [17–22]). While the promise of direct surveys of wildlife from space is an exciting frontier for wildlife biology, the challenges in identifying animals, which are almost always rare and usually only a few pixels in size, remain formidable. In addition to the challenges inherent to classification, the volumes of high-resolution imagery that must be annotated for a comprehensive survey are enormous and require advances in computing, storage, and cyberinfrastructure.

Cetaceans remain a challenging taxon of study given their frequently broad ranges and marine life-history, but their size makes them an attractive target for the use of imagery-based surveys. Previous efforts to locate whales using high-resolution imagery [23–25] have been largely successful yet face challenges in bringing the process to broader spatial or temporal scales given the time required for analysis [25]. Fretwell et al. [24] successfully identified Southern Right whales (*Eubalaena glacialis*) near Peninsula Valdés, Argentina both manually and using supervised and un-supervised classification algorithms in WorldView-2 satellite imagery. This satellite sensor provides 8 spectral bands and a panchromatic band with a maximum resolution of 46 cm per pixel on-nadir. More recently, WorldView-3 imagery has become available, providing a maximum resolution of 31 cm per pixel on-nadir. Cubaynes et al. [25] found that the spectral response of whales declines above the visible red band (630–690 nm). Because both manual and automated methods rely on only the visible bands, automated algorithms are easily validated by visual inspection.

Ocean basins are large, and the potential habitat of cetaceans may encompass vast areas, especially during times of migration. Without knowing where cetaceans are, far more imagery must be considered than can reasonably be annotated manually, particularly if surveys are going to be repeated with any regularity. With this increase in data volume, the only practical solution becomes a workflow that involves a high level of automation to accomplish the otherwise tedious task of manually examining millions of pixels constituting hundreds or thousands of square kilometers. Although contemporary machine learning algorithms have been in use for 20 years or more, their application to ecological datasets have only become commonplace in the past 5–10 years [26]. Machine learning applications involving computer vision are only now gaining traction as a means of managing large volumes of image data that are tedious to analyze manually, such as camera traps, aerial imagery, or time-lapse photography [27]. Such algorithms can automate the process of classifying individual features in images (e.g., [28–29]) and counting or estimating abundance (e.g., [30–31]). Driven by commercial applications, machine learning methods have progressed rapidly in the past decade, with particular interest being paid to deep-learning methods. Deep-learning algorithms have shown promise in the

field of ecology, with applications in acoustic signal detection [32], behavioral predictions [33], and camera trap classification [34], and have been adopted or proposed in numerous other fields such as medicine [35] and traffic management (e.g. [36]).

We present a cetacean survey method, employing a convolutional neural network (CNN) to automate much of the satellite imagery interpretation. Our goal in this initial pilot study was not to develop a fully-automated method, but to identify images with a high probability of containing a whale and thereby minimize the labor required for expert annotation. Here we describe an initial pipeline for whale detection that makes substantial advances toward a fully-automated detection system up to and including the global scale.

## Methods

### Imagery

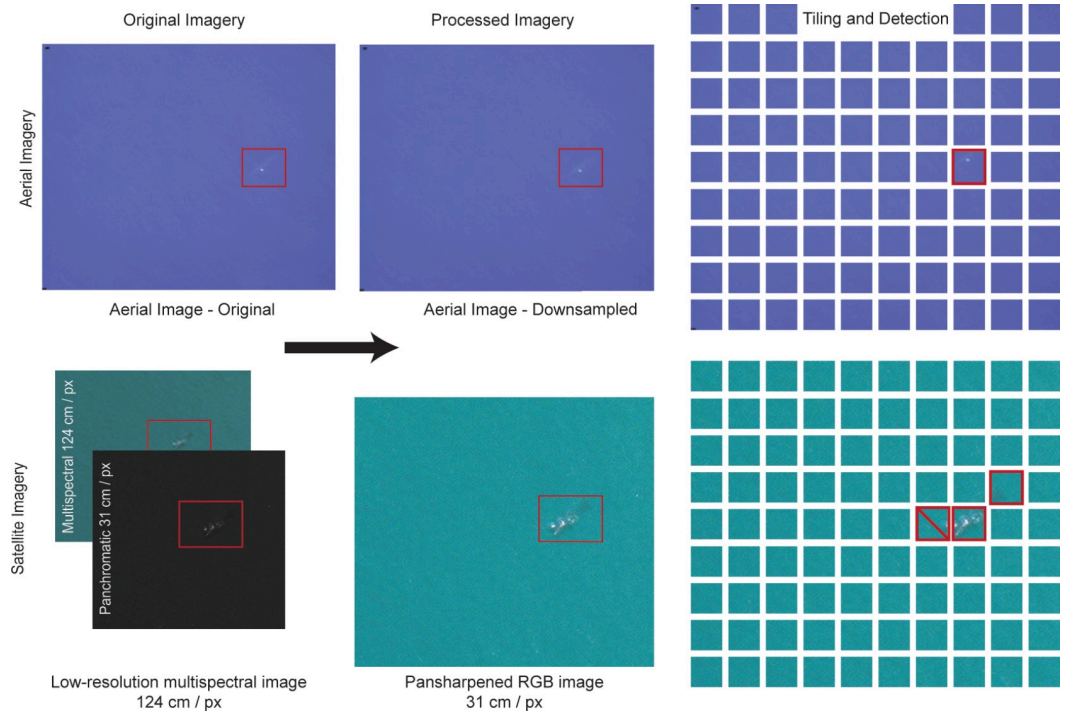
We pooled aerial imagery extracted from high-resolution video footage captured over various water bodies surrounding northern Europe and the United Kingdom by HiDef Aerial Surveying Ltd. to create a training set of whale and water images (Fig 1). The native resolution of the aerial imagery was approximately 2 cm per pixel ground-sample distance, which we down-sampled using a bilinear resampling function in ImageJ [37] to match the 31 cm resolution of Worldview-3 imagery. During the aerial survey, the aircraft flies at 549 m above sea level with a speed of 222 km/h and captures multiple image frames of the same whale using four cameras. The two inner cameras cover a 129 m wide strip and the outer cameras a 143 m wide strip; they are separated by a gap of about 20 m totaling to an effective transect width of 544 m [38]. We elected to retain these duplicate images, as each frame captures the whale at a slightly different angle or in a different body position and is therefore valuable in building the training data set. In total, we had 190 aerial images representing 17 individual minke whales (*Balaenoptera acutorostrata*).

All satellite imagery was collected at a resolution of 31 cm per pixel (on-nadir) from DigitalGlobe's Worldview-3 sensor (DigitalGlobe, Westminster, Colorado; S1 Table). This sensor records imagery in a variety of spectral bands in the visible and near-infrared range. Multi-spectral bands have a lower resolution (124 cm / pixel) than the panchromatic band, which incorporates a broad swath of the visible spectrum and is collected at 31 cm per pixel. To obtain very high-resolution multi-spectral imagery, we pansharpened the lower resolution multi-spectral bands using the higher resolution panchromatic band using the Gram-Schmidt algorithm implemented in ENVI (Exelis Visual Information Solutions, Boulder, Colorado). While pansharpening was conducted using all the available bands, we used only the red (630–690 nm), green (510–580 nm), and blue (450–510 nm) bands to approximate the RGB image captured during aerial survey.

Given that we required satellite imagery that definitely included whales, we surveyed known whale hotspots using Google Earth Pro and located imagery that contained visible Southern Right whales (*Eubalaena australis*) from Peninsula Valdés, Argentina and Humpback whales (*Megaptera novaeangliae*) from Maui, Hawaii (Fig 2). We also acquired cloud-free imagery of these regions based on times of the year where whales would be very likely to be present and active at the surface (S1 Table). We excluded portions of the acquired imagery in which sea conditions prevented manual detection of whales.

### Processing

To prepare imagery for use in model training, we split each image into small  $32 \times 32$ -pixel tiles each measuring  $98.4 \text{ m}^2$  in area (Fig 1; S1 File). Different models require different input sizes and as such, we enlarged the  $32 \times 32$ -pixel tiles to  $224 \times 224$  pixels (S2 File) For both aerial and



**Fig 1. The automated workflow.** Aerial imagery (above) is down-sampled, tiled, and then used to train the model. Satellite imagery (below) is pansharpended and tiled before the model can detect whales. Satellite imagery published under a CC BY license, with permission from the DigitalGlobe Foundation, original copyright 2014.

<https://doi.org/10.1371/journal.pone.0212532.g001>

satellite imagery, tiles were manually separated into whale and water classes. We chose subsets of each satellite image for testing but retained all aerial water tiles for training, resulting in 40,416 aerial tiles (water:  $n = 39,726$ ; whale:  $n = 690$ ) and 40,516 satellite tiles (water:  $n = 40,474$ ; whale:  $n = 42$ ). Of the water satellite tiles, we randomly selected a subset ( $n = 1,390$ ) to reduce testing time.



**Fig 2. Locations of satellite imagery used.** Maui, Hawaii (A) and Peninsula Valdés, Argentina (B) [39].

<https://doi.org/10.1371/journal.pone.0212532.g002>

We selectively removed some aerial tiles prior to creating validation folds. These images contained a miniscule portion of a whale that was recognizable to a human observer only when put into context with the surrounding tiles. Given that these small whales were better captured by other neighboring tiles, we removed them from both training and testing to avoid confusion. Where whales were cleanly bisected by the border between tiles, such that at least 20% of the whale remained in the tile, we retained these images. The same procedure was adopted in with the satellite imagery to avoid false-negatives for whales that would already be captured by the model in another tile, reducing the number of satellite whale tiles used for testing from 42 to 32 (Fig 1).

## Deep learning with convolutional neural networks

CNNs are extensions of traditional neural networks that work by summarizing patterns in images across many “hidden” layers through ‘convolutions’ (i.e., complex data transformations, see [7, 40]). Neural networks, the basis for CNNs, operate by simulating how neurons transmit information through the central nervous system [41–42]. Information is input at a commencing node, which then passes information to several other nodes, which pass on to more nodes, until the output is translated at a terminal node. Each node adds some function to decode the information passed into the commencing node. In CNNs, nodes are organized into “hidden” layers, where each layer applies a different image transformation (or convolution) to information from the previous layer. Transformations could include tasks such as stretching, skewing, splitting, or changing the contrast of images. During training, the training images—or images containing only the target classes, whales and water—are fed to the model so that it can ‘learn,’ i.e. adjust the network’s parameters to minimize the differences between the network’s outputs and the correct labels. The model, once converged, can be validated using a test image subset that was withheld from the training process. This form of machine learning minimizes the need for manually designing a rule-based classification scheme, but limits the interpretability of the model, which acts as a black box. Deep learning is a rapidly-evolving field with new architectures regularly outstripping the performance of previous methods.

## Model training

We trained our model using down-sampled aerial imagery. We separated 90% of aerial imagery for use in training, and 10% for use in validating the trained model, repeating this process to create a 10-fold validation system wherein we iteratively trained and tested on each fold to verify that no set of images was having an undue influence on the model (S1 Fig). We tested our model’s classification performance by applying it to very high-resolution satellite imagery. Manually-annotated satellite imagery, in which whales and empty ocean were identified by expert annotators, was used as the standard against which the CNN’s performance was compared. It is important to note that our model was naive to ‘real’ satellite imagery and was applied without further refinement following its development using the down-sampled aerial training and testing dataset.

We implemented our CNN using the Pytorch framework [43], which makes it easy to implement, train, and adapt a model, and tested three different architectures: ResNet (using 18-, 34-, and 152-layer models) [44] and DenseNet [45]. These models are all widely-used and have performed well in various competitions. Each network takes as input a small image and outputs a vector of two elements, which represent the probability of the input image containing or not containing a whale, respectively. All code (S1–S6 Files), required software packages (S2 Table), and details on hardware used (S7 File) are included in the supplementary materials. Each model was pre-trained on the ImageNet dataset [46] consisting of 1.28 million training



images of 1,000 different classes (e.g. ‘house,’ ‘spider,’ ‘fire’). We modify the last layer of this model to train with our data, i.e. having only 2 classes as opposed to 1,000. Pre-training the model on ImageNet has been shown to reduce overfitting and training time [46].

Whales are inherently rare in imagery and the training architecture takes only small batches ( $n = 4$  to  $n = 32$ ) of images at a time, so we risk having the model examine only images of water much of the time. As this may impair the ability of the model to learn, we implemented a weighted random sampler that increases the probability that a whale image will appear in any given batch in proportion to the number of whale images in the overall training set (S4 and S5 Files). Users can set several hyperparameters and we experimented primarily with the learning rate which governs how new information is weighted against older information. Higher learning rates down-weight older information relative to new information; if the learning rate is too high, the model will disregard previous whales it has seen in favor of the characteristics of the most recent whale it has encountered. Conversely, if the learning rate is too low, and the model is reluctant to incorporate new information. We used the set of images that were withheld from training to test the performance of each model. False positives and false negatives for each epoch were used to further tune model parameters and retrain models.

To demonstrate the value of the CNN approach, we also trained a ridge regression model ( $\alpha = 1$ ) and C-Support Vector Classifier model (C-SVC;  $C = 1$ ) [47] using a reduced training set (retaining all whale images but randomly selecting an equal number of water images) and the same testing set, and implemented through Scikit-learn ([48]; S8 File). These methods, in contrast to the CNN approach, require an additional feature-extraction step implemented using a histogram of oriented gradients approach.

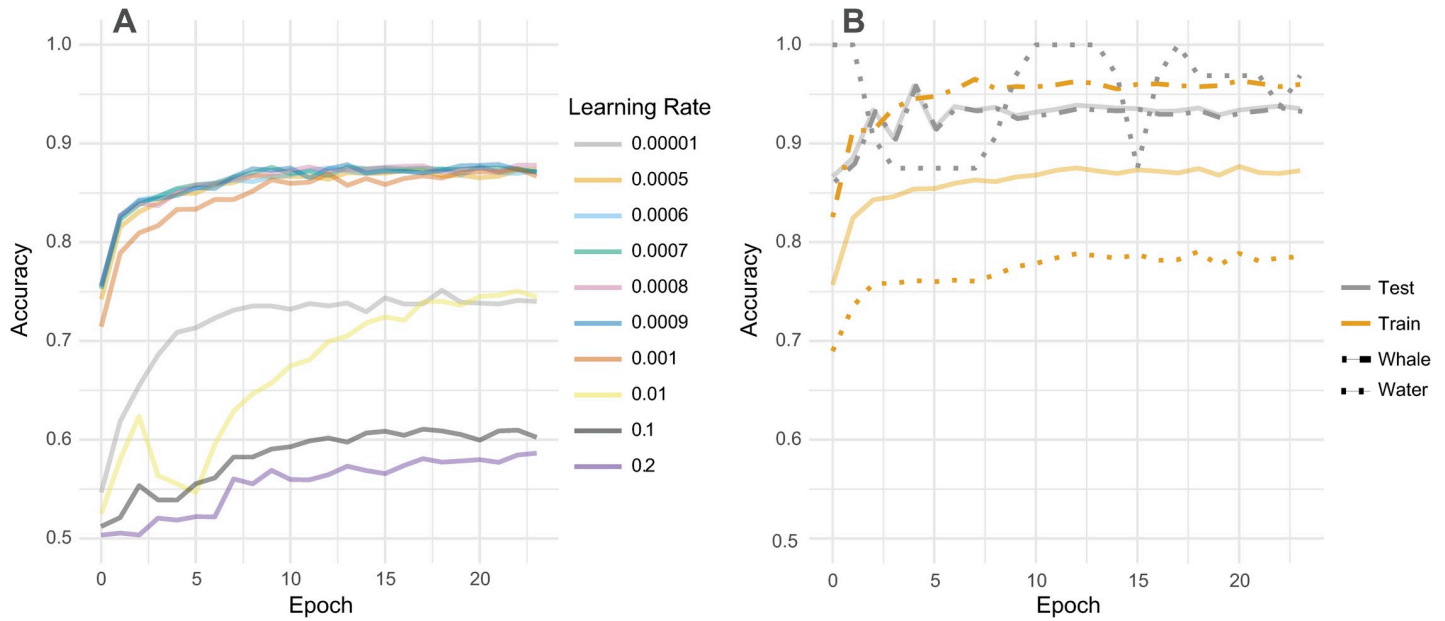
## Results

Image tiling took approximately 15 seconds per  $\text{km}^2$  though the precise timing is sensitive to the specific characteristics of the computing resources available. Model training time varied among CNN model architectures from approximately one hour (ResNet-18) to nearly 7 hours (DenseNet) for the full training set on our hardware (S7 File). We trained all models for 24 epochs and tested the model weights of each epoch against our test set of satellite images. Our best model trained for 9 epochs and used a learning rate set at 0.0006, with a step size (a parameter that allows the learning rate to decay after a certain number of epochs) of seven epochs, and a momentum of 0.9 (Table 1). We found that a higher learning rate often led to little learning, wherein the model would perform only slightly better than random chance on the training dataset (Fig 3). On the other hand, a model with a very low learning rate of 0.00001

**Table 1. Results of variation in learning rate (LR) for ResNet-152.**

Learning rate	Precision	Recall
LR = 0.00001	1.000	0.722
LR = 0.0005	1.000	0.797
<b>LR = 0.0006</b>	<b>1.000</b>	<b>0.937</b>
LR = 0.0007	1.000	0.860
LR = 0.0008	1.000	0.908
LR = 0.0009	1.000	0.895
LR = 0.001	1.000	0.915
LR = 0.01	0.996	0.707
LR = 0.1	0.987	0.898
LR = 0.2	0.986	0.950

<https://doi.org/10.1371/journal.pone.0212532.t001>



**Fig 3. Effect of learning rate on performance.** (A) Each model was trained using several different learning rates, which affected their performance on the training dataset (S9 File). (B) The best model weights came from the 12<sup>th</sup> epoch of a ResNet-152 model with a learning rate of 0.0006 (S10 File). In both training and testing phases, solid lines show overall performance, while dashed and dotted lines show the accuracy of the individual classes. With a step-size of 7, the learning rate begins to decay at the 7<sup>th</sup> epoch, and the accuracy begins to level out.

<https://doi.org/10.1371/journal.pone.0212532.g003>

learned very slowly During the training phase of the ten-fold validation, all folds trained along a similar trajectory.

The standard measurement of performance in this case is precision (the percent of positives [model-classified as water] that are true positives [manually-labeled as water]), and recall (the percent of manually-labeled water images that were found by the model). Precision was closely consistent among folds, ranging from 0.99 to 1.00 (that is, nearly everything classified as water was actually water such that few or no whales were misclassified), with higher variation in recall (0.33 to 1.00). This suggests that there was some variation among the training images in the folds that was affecting model performance, though most of this variation falls in only folds 1 and 10 (Table 2; S1 Fig). In particular, fold 10 had very poor performance, likely due to the fact that the test images in this fold had much rougher sea conditions than in the other scenes.

**Table 2. Results at the final epoch of 10-fold validation.**

Aerial test fold	N water training	N whale training	N water test	N whale test	Precision	Recall
1	11,076	207	1,230	23	0.996	0.616
2	11,076	207	1,230	23	1.000	0.988
3	11,076	207	1,230	23	0.993	0.994
4	11,076	208	1,230	23	0.999	0.999
5	11,076	208	1,230	23	0.999	0.989
6	11,076	207	1,230	23	1.000	0.996
7	11,076	207	1,230	23	1.000	0.998
8	11,076	207	1,230	23	1.000	1.000
9	11,076	207	1,230	23	0.955	1.000
10	12,306	206	1,236	24	0.998	0.322

<https://doi.org/10.1371/journal.pone.0212532.t002>

Table 3. Performance of different model types.

Model	Precision	Recall	F1 Score	Epoch	LR
ResNet-18	1.000	0.932	0.965	24	0.0009
ResNet-34	1.000	0.932	0.965	20	0.0009
ResNet-152	<b>1.000</b>	<b>0.937</b>	<b>0.968</b>	<b>9</b>	<b>0.0006</b>
DenseNet	0.999	0.922	0.959	5	0.001

<https://doi.org/10.1371/journal.pone.0212532.t003>

Given that this fold was trained only on calmer conditions, it is not surprising that the edges and contrast of these rougher images could be mis-classified as whales (S2 Fig).

All CNN model architectures succeeded in finding all or nearly all whales in our test set, and correctly classified nearly all water images, with the best model performance resulting in an F1 score,  $\frac{2 * precision * recall}{precision + recall}$ , of 0.968 (Table 3; Fig 4). By contrast, the more traditional classification methods (ridge regression and C-SVC) performed more poorly than any of the CNN models, finding only 88% of the whales in the test set (Ridge regression: precision = 0.996, recall = 0.678, F1 = 0.807; C-SVC: precision = 0.995, recall = 0.632, F1 = 0.773). Modifying the parameter  $\alpha$  in ridge regression determines the weight given to the residual sum of squares and the sum of square coefficients. The outcome for  $\alpha = 0$  was F1 = 0.827 versus 0.804 for  $\alpha =$

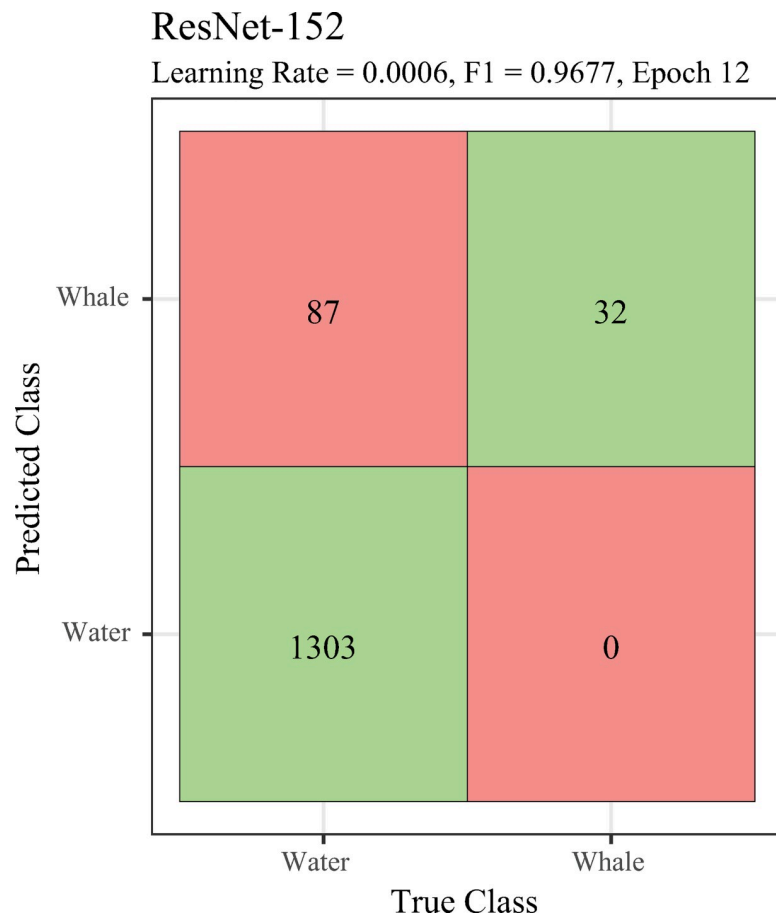


Fig 4. Model performance. Confusion matrix for the best model—ResNet-152. Full table of performance for all models available in S11 File.

<https://doi.org/10.1371/journal.pone.0212532.g004>



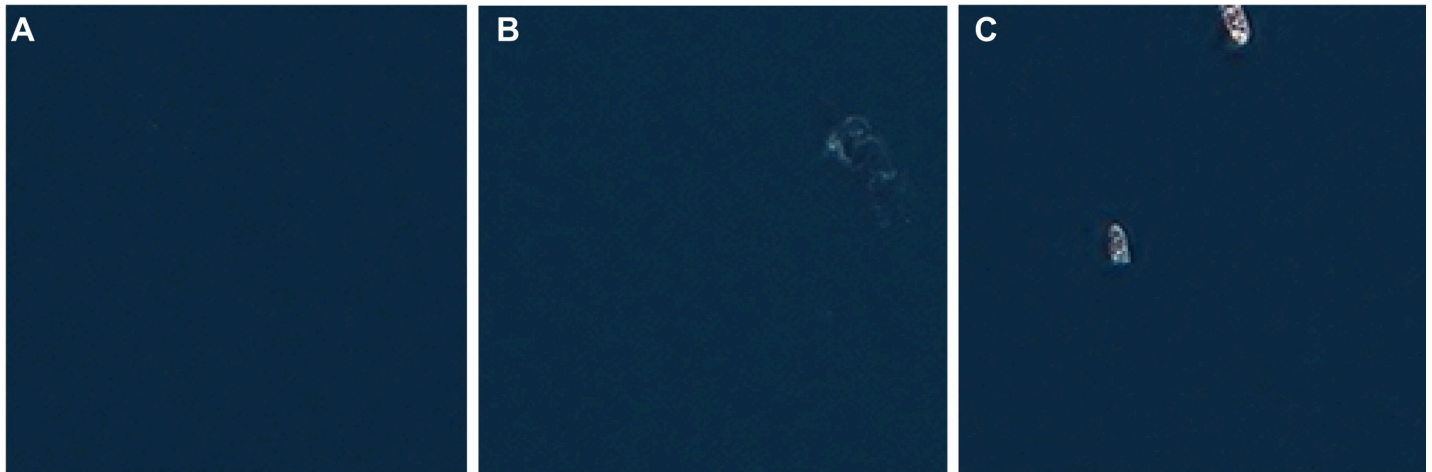
10,000, suggesting that for this problem, ridge regression is little better than a simple linear regression. For C-SVC,  $C$  is a penalty term that controls the prioritization of classification accuracy versus smooth boundaries. Similar to ridge regression, tuning this parameter did not significantly improve the model ( $C = 0.001$ ,  $F1 = 0.773$ ;  $C = 10,000$ ,  $F1 = 0.799$ ).

## Discussion

Here we describe a functioning pipeline for identifying whales in high-resolution satellite imagery that can be immediately employed to reduce the time required to complete large-extent surveys. All ResNet versions performed well, correctly classifying all whales and at this stage we recommend ResNet-152 for its high accuracy, correctly classifying all whales and mis-classifying only 87 of 1390 water images in WorldView-3 imagery for a false positive rate of about 6.1% (Table 3; Fig 4). DenseNet also performed well but did not match the success of ResNet. Neither C-SVC nor ridge regression matched the performance of any of the CNN models, likely reflecting the subtle appearance of whales in imagery and the degree of variation among whales and among scenes involving different water conditions and suggesting that neither is well-suited to this task. In this case the F1 score demonstrates the quality of the ResNet-152 model, though these scores can be misleading. In our case it is more important to maximize precision than recall, as false positives can easily be thrown out, but examining all false negatives requires the same amount of time as manually annotating, defeating the purpose of the automated approach. Importantly, the assignment of one class or another as “positive” or “negative” is arbitrary. Our model considered water as the positive case; were it the other way around, it would be more important to maximize recall.

We have deployed multiple model architectures here and received promising results with several of them, yet future development in deep learning will likely outstrip their performance. We have tuned our models to our particular problem and dataset, but the optimal parameters for our dataset are not universal. For example, ResNet-18 required a full 24 epochs of training to reach model weights that performed well on satellite imagery. ResNet-152, a “deeper” and “wider” network, arrived at its best weights after only 9 epochs (Table 3). In this case, both were trained with the same set of experimental learning rates, but this and other parameters could be tuned to allow for longer or shorter training. In the case of ResNet-152, the model weights at the 24<sup>th</sup> epoch demonstrated that the model had overfit the training data slightly, and this can easily change with a different training set, especially one of different size.

Like most machine-learning applications, the model could be iteratively improved with the addition of correctly classified and verified whales from future imagery, and our current classification accuracy therefore represents a lower bound on the potential for satellite imagery to aid in cetacean surveys. We believe that this method is an improvement over the previous classification methods employed by Fretwell et al. [24] which, while successful, will likely be sensitive to differences in ocean color and turbidity and less robust to the size of different species. Our method differs from Fretwell et al. [24] in several respects, not only in regard to the classification method but also with respect to the underlying data used (that is, red, green, and blue bands only), and one avenue for further research may be to explore the benefit of using additional spectral bands for classification. Despite the demonstrated feasibility of automated classification of whales in satellite imagery, barriers remain to broad adoption. Most significant is the paucity of open-water imagery available in DigitalGlobe’s archive (and the lack of similar-resolution sensors from other providers), which reflects that imagery is not collected continuously but is instead targeted within high-demand regions or in response to specific orders from customers. Hopefully, more interest in using satellite imagery for marine mammal surveys will facilitate the expansion of open-water imagery available within the catalog.



**Fig 5. Example images.** Open water (A), Southern right whale (B), whale-watching boat (C). Satellite imagery published under a CC BY license, with permission from the DigitalGlobe Foundation, original copyright 2014.

<https://doi.org/10.1371/journal.pone.0212532.g005>

The current pricing structure for very high-resolution imagery would likely prevent many research applications from pursuing projects at basin-wide scales, but alternative pricing for non-profit organizations and education users is available. While it is difficult to estimate the cost of field surveys given the differing logistics based on time and region, Abileah [49] suggests costs should be similar to aerial surveys and a substantial savings in more remote areas. Satellite tasking logistics can make imagery acquisition in some locations (e.g., high latitudes) challenging, though plans for larger satellite constellations will ameliorate many of these limitations over time. Encouraging the collection of open-water imagery in areas of interest and in areas of low competition is the first step in moving imagery-based methods to broad applicability. In the meantime, the pooling of aerial photographs with known ground-sample distance by different research groups could result in a more robust training image set, and users with large catalogs of aerial imagery for their specific taxa and regions could create bespoke local training sets to better classify cetaceans in their region of interest.

Given that our model is trained exclusively on minke whales, the smallest of the baleen whales, including more aerial photography of larger whale species would likely further improve performance. That said, it performed surprisingly well on whales that can reach double the size of a minke whale. The code could easily be modified to create overlapping tiles, which would eliminate problems arising from whales bisected by neighboring tiles (S1 File). The addition of further classes representing objects such as boats, large ships, land, and rocks at the surface in the imagery would also help minimize the number of false positives in a cetacean survey (Fig 5). A greater number of classes would allow the model to more accurately classify objects that don't fit neatly into the water or whale category; currently those objects are forced into one of the existing categories even if they are a poor visual match. Such classes could be fine-tuned to the application at hand with training images added for the particular conditions found in a region, such as peculiar boat shapes or floating rafts of detritus. Further classification to the species level is theoretically possible but only with a much more robust training set. Work on manual species classification from satellite imagery shows promise but also indicates that some species are more readily identifiable than others [25].

Limitations inherent to this method are not dissimilar to those faced by any other survey method. Challenging sea-state conditions are common to boat- and aerial-based surveying [50–53] because waves and sea spray create a lot of “noise” which makes it difficult to separate

a whale from the surrounding water. We did not attempt to use satellite imagery to locate whales in choppy water. It is worth noting, however, that aerial or satellite methods would likely have more success than surface-level observations as the orthogonal view allows the observer to see at least partially through the water. While we have not developed a definitive threshold for sea state, we expect that observations above Beaufort-4 observations would be difficult given widespread whitecaps at the surface. Below Beaufort-4, the size of “noisy” elements on the water below Beaufort-4 are likely to be in the range of a single pixel (31 x 31 cm) and discrimination of whales feasible. Cloud cover is also a controlling factor, and future applications will need to pair this detection pipeline with appropriate statistical models for non-detection [54].

Aside from environmental challenges, there are several satellites currently in orbit that could be used for cetacean surveying, such as previous iterations of WorldView and Pleiades. While this trained model may be robust to differences in spatial resolution among the various sensors available, we did not test imagery from other sensors. Sensor-specific models could easily be trained, as the aerial imagery can be down-sampled to any resolution desired. Scaling this method to process larger volumes of imagery will be manageable for an individual user for small areas, but once the spatial and temporal scope increases, tailored cyberinfrastructure (such as is underway with the ICEBERG project; <https://iceberg-project.github.io>) will be required to handle both the storage and transmission of imagery to a computing cluster [55].

This method could be used to improve cetacean research in several different ways. It provides a means of viewing and monitoring areas that are far from ports or are hazardous to access, such as polar regions, remote island chains, or open ocean. Moreover, it provides the potential to monitor these areas at a daily time scale, cloud-cover permitting. With enough imagery, it could be used to monitor the arrival of migrating species or examine fine-scale changes in foraging activity. Long-term studies on whale feeding and breeding grounds have provided critical information on the ecology and behavior of these animals but are poorly suited to answer basin-scale questions for species that range widely both within and among seasons. The arrival time of migrating whales at traditional feeding grounds, for example, has been used to understand links between habitat use and local environmental conditions [56], but is unable to illuminate the existence of unmonitored areas that may serve as alternative feeding grounds. With the ability to rapidly and automatically detect whales in satellite imagery, boat or aerial surveys become valuable as ground-truthing rather than as the sole source of data on whale abundance and distribution, and researchers intent on instrumenting individuals or collecting individual-level data may be able to more accurately target their effort, saving time and expense. While far from a total replacement for other survey modalities, this method has promise to improve current survey methodology for large whales, increase the temporal resolution of surveys, expand the ocean surface area surveyed, minimize human risk, and increase the rate of data acquisition.

## Supporting information

**S1 Table. Satellite imagery.** We acquired imagery from Digital Globe’s WorldView-3 sensor via the Digital Globe Foundation. See <https://discover.digitalglobe.com/> for details on individual scenes and a preview.

(PDF)

**S2 Table. Python packages.** The code requires packages for Python 3 to be pre-installed.

(PDF)

**S1 File. Image tiling code.**

(TXT)

**S2 File. Image resizing code.**  
(TXT)

**S3 File. Model testing code.**  
(TXT)

**S4 File. Code utilities.**  
(TXT)

**S5 File. Model training code.**  
(TXT)

**S6 File. Model utility.**  
(TXT)

**S7 File. Details on hardware.**  
(PDF)

**S8 File. Ridge regression/SVM code.**  
(TXT)

**S9 File. Training accuracy.** Accuracy and loss at different learning rates, used to create [Fig 3A](#).  
(CSV)

**S10 File. Best model results.** Results at the training and testing phase for a ResNet-152 model. Used to create [Fig 3B](#).  
(CSV)

**S11 File. Model outcomes.** Classification outcomes for each model tested.  
(CSV)

**S1 Fig. Accuracy, loss, and testing results on validation folds.** Training and testing results for 10-fold validation: accuracy and loss for each fold during the training process (A). The precision and recall for each fold (B).  
(PDF)

**S2 Fig. Sea conditions potentially affecting 10-fold validation.**  
(PDF)

## Acknowledgments

The work is part of the kick-start activity SPACEWHALE, funded by the European Space Agency to BioConsult SH. The authors gratefully acknowledge an imagery grant to AB from the Digital Globe Foundation, computational time from the Seawulf Cluster at the Institute of Advanced Computational Science (NSF grant #1531492), aerial imagery from HiDef Aerial Surveying Ltd., and enlightening discussion with Bento Gonçalves, Peter Fretwell, and Ellie Bowler.

## Author Contributions

**Conceptualization:** Alex Borowicz, Grant Humphries, Georg Nehls, Caroline Höschle, Vladislav Kosarev, Heather J. Lynch.

**Data curation:** Alex Borowicz, Grant Humphries.

**Formal analysis:** Alex Borowicz, Grant Humphries.

**Funding acquisition:** Georg Nehls.

**Investigation:** Alex Borowicz, Grant Humphries.

**Methodology:** Alex Borowicz, Hieu Le, Grant Humphries, Vladislav Kosarev, Heather J. Lynch.

**Project administration:** Alex Borowicz, Georg Nehls, Caroline Höschle.

**Resources:** Grant Humphries, Georg Nehls, Heather J. Lynch.

**Software:** Alex Borowicz, Hieu Le, Grant Humphries.

**Supervision:** Georg Nehls, Heather J. Lynch.

**Validation:** Alex Borowicz.

**Visualization:** Alex Borowicz.

**Writing – original draft:** Alex Borowicz.

**Writing – review & editing:** Hieu Le, Grant Humphries, Georg Nehls, Caroline Höschle, Vladislav Kosarev, Heather J. Lynch.

## References

1. Hamazaki T. Spatiotemporal prediction models of cetacean habitats in the mid-western North Atlantic Ocean (from Cape Hatteras, North Carolina, U.S.A. to Nova Scotia, Canada). *Mar Mamm Sci.* 2002; 18: 920–939.
2. Yen PW, Sydeman WJ, Hyrenbach KD. Marine bird and cetacean associations with bathymetric habitats and shallow water topographies: implications for trophic transfer and conservation. *J Mar Syst.* 2004; 50: 70–99.
3. Moore SE, Waite JM, Mazzuca LL, Hobbs RC. Mysticete whale abundance and observations of prey associations on the central Bering Sea shelf. *J Cetacean Res Manag.* 2010; 2: 227–234.
4. Bedriñara-Romano L, Hucke-Gaete R, Viddi FA, Morales J, Williams R, Ashe E, et al. Integrating multiple data sources for assessing blue whale abundance and distribution in Chilean Northern Patagonia. *Divers Distrib.* 2018; 24: 991–1004.
5. Kaschner K, Quick NJ, Jewell R, Williams R, Harris CM. Global coverage of cetacean line-transect surveys: status quo, data gaps and future challenges. *PLoS ONE.* 2012; 7(9): e44075. <https://doi.org/10.1371/journal.pone.0044075> PMID: 22984461
6. Lesage V, Gavrilchuk K, Andrews RD, Sears R. Foraging areas, migratory movements and winter destinations of blue whales from the western North Atlantic. *Endanger Species Res.* 2017; 34: 27–43.
7. Weinstein BG, Friedlaender AS. Dynamic foraging of a top predator in a seasonal polar marine environment. *Oecologia.* 2017; 185: 427–435. <https://doi.org/10.1007/s00442-017-3949-6> PMID: 28914358
8. Bodey TW, Cleasby IR, Bell F, Parr N, Schultz A, Votier SC, et al. A phylogenetically controlled meta-analysis of biologging device effects on birds: Deleterious effects and a call for more standardized reporting of study data. *Methods Ecol Evol.* 2018; 9: 946–955.
9. Grecian WJ, Lane JV, Michelot T, Wade HM, Hamer KC. Understanding the ontogeny of foraging behaviour: insights from combining marine predator bio-logging with satellite-derived oceanography in hidden Markov models. *J R Soc Interface.* 2018; 15(143): 20180084 <https://doi.org/10.1098/rsif.2018.0084> PMID: 29875281
10. McKinnon EA, Love OP. Ten years tracking the migrations of small landbirds: Lessons learned in the golden age of bio-logging. *The Auk.* 2018; 135: 834–856.
11. Kot CY, Fujioka E, Hazen LJ, Best BD, Read AJ, Halpin PN. Spatio-temporal gap analysis of OBIS-SEAMAP project data: Assessment and way forward. *PLoS ONE.* 2010; 5(9): e12990. <https://doi.org/10.1371/journal.pone.0012990> PMID: 20886047
12. Smith TD, Allen J, Clapham PJ, Hammond PS, Katona S, Larsen F, et al. An ocean-basin-wide mark-recapture study of the North Atlantic humpback whale (*Megaptera novaeangliae*). *Mar Mamm Sci.* 1999; 15(1): 1–32.
13. Branch T. Abundance of Antarctic blue whales south of 60°S from three complete circumpolar sets of surveys. *J Cetacean Res Manag.* 2007; 9: 253–262.

14. Branch T. Humpback whale abundance south of 60°S from three complete circumpolar sets of surveys. *J Cetacean Res Manag.* 2011; 3: 53–69.
15. Hammond PS, Lacey C, Gilles A, Viquerat S, Boerjesson P, Herr H, et al. Estimates of cetacean abundance in European Atlantic waters in summer 2016 from the SCANS-III aerial and shipboard surveys. Wageningen Marine Research. 2017.
16. Branch T, Stafford KM, Palacios DM, Allison C, Bannister JL, Burton CLK, et al. Past and present distribution, densities, and movements of blue whales *Balaenoptera musculus* in the Southern Hemisphere and northern Indian Ocean. *Mamm Rev.* 2007; 37: 116–175.
17. LaRue MA, Rotella JJ, Garrott RA, Siniif DB, Ainley DG, Stauffer GE, et al. Satellite imagery can be used to detect variations in abundance of Weddell seals (*Leptonychotes weddelli*) in Erebus Bay, Antarctica. *Polar Biol.* 2011; 34: 1727–1737.
18. Fretwell PT, LaRue MA, Morin P, Kooyman GL, Wienecke B, Ratcliffe N, et al. An emperor penguin population estimate: The first global, synoptic survey of a species from space. *PLoS ONE.* 2012; 7(4): e33751. <https://doi.org/10.1371/journal.pone.0033751> PMID: 22514609
19. Platonov NG, Mordvintsev IN, Rozhnov VV. The possibility of using high resolution satellite images for detection of marine mammals. *Biol Bull.* 2013; 40: 197–205.
20. Lynch HG, LaRue MA. First global census of the Adélie Penguin. *The Auk.* 2014; 131:457–466.
21. Fretwell PT, Scofield P, Phillips RA. Using super-high resolution satellite imagery to census threatened albatrosses. *Ibis.* 2017; 159: 481–490.
22. Seymour AC, Dale J, Hammill M, Halpin PN, Johnston DW. Automated detection and enumeration of marine wildlife using unmanned aircraft systems (UAS) and thermal imagery. *Sci Rep.* 2017; 7: 45127. <https://doi.org/10.1038/srep45127> PMID: 28338047
23. Abileah R. Marine mammal census using space satellite imagery. *US Navy J Underw Acoust.* 2002; 52(3): 709–724.
24. Fretwell PT, Staniland IJ, Forcada J. Whales from space: Counting Southern Right whales by satellite. *PLoS ONE.* 2014; 9(2): e88655. <https://doi.org/10.1371/journal.pone.0088655> PMID: 24533131
25. Cubaynes HC, Fretwell PT, Bamford C, Gerrish L, Jackson JA. Whales from space: Four mysticetes species described using new VHR satellite imagery. *Mar Mamm Sci.* Forthcoming 2019.
26. Humphries G, Magness DR, Huettmann F, editors. *Machine learning for ecology and sustainable natural resource management.* 1<sup>st</sup> ed. Cham, Switzerland: Springer Nature; 2018.
27. Weinstein B. A computer vision for animal ecology. *J Anim Ecol.* 2017; 87: 1–13.
28. Zhang Z, He Z, Cao G, Cao W. Animal detection from highly cluttered natural scenes using spatiotemporal object region proposals and patch verification. *IEEE Transa Multimedia.* 2016; 18(10): 2079–2092.
29. Willi M, Pitman RT, Cardoso AW, Locke C, Swanson A, Boyer A, et al. Identifying animal species in camera trap images using deep learning and citizen science. *Methods Ecol Evol.* Forthcoming 2019.
30. Barber-Meyers SM, Kooyman GL, Ponganis PJ. Estimating the relative abundance of emperor penguins at inaccessible colonies using satellite imagery. *Polar Biol.* 2007; 30: 1565–1570.
31. Seiferling I, Naik N, Ratti C, Proulx R. Green streets—Quantifying and mapping urban trees with street-level imagery and computer vision. *Landsc Urban Plan.* 2017; 165: 93–101.
32. Aodha OM, Gibb R, Barlow KE, Browning E, Firman M, Freeman R, et al. Bat detective—Deep learning tools for bat acoustic signal detection. *PLoS Comput Biol.* 2018; 14: e10059995.
33. Browning E, Bolton M, Owen E, Shoji A, Guilford T, Freeman R. Predicting animal behaviour using deep learning: GPS data alone accurately predict diving in seabirds. *Methods Ecol Evol.* 2018; 9: 681–692.
34. Nourouzzadeh MS, Nguyen A, Kosmala M, Swanson A, Palmer MS, Packer C, et al. Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proc Natl Acad Sci U S A.* 2018; 115(25): E5716–E5725. <https://doi.org/10.1073/pnas.1719367115> PMID: 29871948
35. Saltz J, Gupta R, Hou L, Kurc T, Singh P, Nguyen V et al. Spatial organization and molecular correlation of tumor-infiltrating lymphocytes using deep learning on pathology images. *Cell Rep.* 2018; 23(1): 181–193. <https://doi.org/10.1016/j.celrep.2018.03.086> PMID: 29617659
36. Lv Y, Duan Y, Kang W, Li Z, Wang F. Traffic flow prediction with big data: A deep learning approach. *IEEE trans Intell Transp Syst.* 2015; 16: 865–873.
37. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods.* 2012; 9: 671–675. <https://doi.org/10.1038/nmeth.2089> PMID: 22930834
38. Weiß F, Büttger H, Baer J, Welcker J, Nehls G. Erfassung von Seevögeln und Meeressäugertieren mit dem HiDef-Kamerasystem aus der Luft. *Seevögel.* 2016; 37(2): 14–21.



39. Tapiquén CE. South America [shapefile]. Porlamar, Venezuela: Orogénesis Soluciones Geográficas; 2015.
40. LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proc IEEE*. 1998; 86(11): 2278–2324.
41. Rosenblatt F. Principles of neurodynamics: Perceptrons and the theory of brain mechanisms. Washington: Spartan Books; 1962.
42. Rumelhart DE, Hinton GE, Williams RJ. Learning internal representations by error propagation. In: Parallel distributed processing: Explorations in the microstructure of cognition Volume I: Foundation. Cambridge, Mass: MIT Press; 1986.
43. Paszke A, Gross S, Chintala S, Chanan G, Yang E, DeVito Z, et al. Automatic differentiation in pytorch. In: NIPS 2017 Autodiff Workshop: The Future of Gradient-based Machine Learning Software and Techniques, 2017 Dec 9; Long Beach, CA, US.
44. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *CVPR 2016: 2016 IEEE Conference on Computer Vision and Pattern Recognition*; 2016 Jun 26-Jul 1; Las Vegas, USA. 770–778.
45. Huang G, Liu Z, van der Maaten L, Weinberger KQ. Densely connected convolutional networks. *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit*. 2017; 4700–4708.
46. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. ImageNet large scale visual recognition challenge. *Int J Comput Vis*. 2015; 115: 211–252.
47. Chang C, Lin C. LIBSVM: A library for support vector machines. *ACM Trans Intell Syst Technol*. 2011, 2: 27–2:27–27.
48. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. *J Mach Learn Res*. 2011, 12: 2825:2830.
49. Abileah R. Marine mammal census using space satellite imagery. *US Navy J Underwater Acoust*. 2002; 52: 709–724.
50. Barlow J, Gerrodette T, Forcada J. Factors affecting perpendicular sighting distances on shipboard line-transect surveys for cetaceans. *J Cetacean Res Manag*. 2001; 3: 201–212.
51. DeMaster DP, Lowry LF, Frost KJ, Bengtson RA. The effect of sea state on estimates of abundance for beluga whales (*Delphinapterus leucas*) in Norton Sound, Alaska. *Fish Bull*. 2001; 99: 197–201.
52. Teilmann J. Influence of sea state on density estimates of harbor porpoises (*Phocoena phocoena*). *J Cetacean Res Manag*. 2003; 5: 85–92.
53. Hodgson A, Kelly N, Peel D. Unmanned aerial vehicles (UAVs) for surveying marine fauna: a dugong case study. *PLoS ONE*. 2013; 8(11): e79556. <https://doi.org/10.1371/journal.pone.0079556> PMID: 24223967
54. Kéry M, Schmidt BR. Imperfect detection and its consequences for monitoring for conservation. *Community Ecol*. 2008; 9: 207–216.
55. Iceberg-project.github.io [Internet]. Imagery cyber-infrastructure and extensible building blocks to enhance geosciences research; c2019 [cited 2019 May 9]. Available from: <https://iceberg-project.github.io/>.
56. Ramp C, Delarue J, Palsbøll PJ, Sears R, Hammond PS. Adapting to a warmer ocean—seasonal shift of baleen whale movements over three decades. *PLoS ONE*. 2015; 10(3): e0121374. <https://doi.org/10.1371/journal.pone.0121374> PMID: 25785462