Automatic whale counting in satellite images with deep learning

Emilio Guirado^{1,2}, Siham Tabik¹, Marga L. Rivas³, Domingo Alcaraz-Segura^{2,4}, Francisco Herrera¹

¹ Andalusian Research Institute in Data Science and Computational Intelligence, University of Granada, 18071, Granada, Spain.

² Andalusian Center for the Assessment and Monitoring of Global Change, University of Almería, 04120, Almería, Spain.

³ Department of Biology, University Marta Abreu of Las Villas, 50100, Santa Clara, Cuba.

⁴ iecolab. Interuniversity Institute for Earth System Research, University of Granada, 18071, Granada, Spain.

Abstract

Despite their interest and threat status, the number of whales in world's oceans remains highly uncertain. Whales detection is normally carried out from costly sighting surveys, acoustic surveys or through high-resolution orthoimages. Since deep convolutional neural networks (CNNs) achieve great performance in object-recognition in images, here we propose a robust and generalizable CNN-based system for automatically detecting and counting whales from space based on open data and tools. A test of the system on Google Earth images in ten global whale-watching hotspots achieved a performance (F1-measure) of 84% in detecting and 97% in counting 80 whales. Applying this cost-effective method worldwide could facilitate the assessment of whale populations to guide conservation actions. Free and global access to high-resolution imagery for conservation purposes would boost this process.

Keywords: deep learning | deep neural networks | artificial intelligence | remote sensing | wildlife ecology

Main

Whales importance, threats and uncertainties

Whales, which comprise some of the largest animals that have ever existed, have always thrilled humans (1, 2). Whales had and keep an enormous economic and societal value (3). More than 13 million whale-watchers were registered in 2008 across 119 countries, generating a global economic activity of US\$ 2.1 billion (4). Since whales generally are long-living species at high trophic-levels, they play an essential role to structure marine food webs, and to maintain ecosystem functions and services (5,6). In the past, commercial whaling depleted whale populations from 66% to 90% from their original numbers, which subsequently caused alterations in marine biodiversity and functions (7). To prevent whales from extinction, the signatories of the International Convention for the Regulation of Whaling limited whale hunting to scientific or aboriginal actions since 1946 (8). Even though, to date, there still exist a great uncertainty around the number of whales in the oceans and the viability of their populations (9). In late 2017, the Species Red List of the International Union for Conservation of Nature (IUCN) (10) reported that 22% of 89 evaluated cetacean species were classified as threatened, whereas almost 50% species could not be evaluated due to the lack of data. Hence, a more accurate estimation of whale distribution and population sizes is essential to warrant cetacean conservation (11).

Classical whale-counting methods

The process of identifying and estimating the number of cetaceans is normally carried out 1) *in situ*, from ships, planes or ground stations (12), by using visual surveys, acoustic methods (13), or a combination of both (14); or 2) *ex situ*, by using satellite tracking (15, 16) or photo-interpretation or classical image classification techniques on aerial or satellite

images (17-21). However, these methods are costly, not robust against scenario changes (e.g., different regions or atmospheric conditions), not generalizable to a massive set of images, and often require handcrafted features (22). Indeed, biodiversity conservation would certainly benefit from robust and automatic systems to assess species distributions, abundances and movements from space (23, 24).

Could deep learning help in whale counting?

Deep learning methods, particularly Convolutional Neural Networks (CNNs), could help in this sense since CNNs are already outperforming humans in visual tasks such as image classification and object detection (25). CNN models have the capacity to automatically learn the distinctive features of different object classes from a large number of annotated images to later make correct predictions on new images (26). Despite building large training datasets is costly, the learning of CNNs on small datasets can be boosted by data-augmentation, which consists of increasing the volume of the training dataset artificially, and by transfer learning, which consists of starting the learning of the network from a prior knowledge rather from scratch (27).

Identifying whales from satellite images using CNNs at a global scale is very challenging for several reasons: 1) comprehensive datasets of whale orthoimages to train CNNs do not exist yet; 2) very high-resolution images are expensive and relatively scarce in the marine environment; 3) whales could potentially be confused with other objects such as boats, rocks, waves, or foam; 4) whale postures or swimming movements captured in a snapshot are quite variable since different parts of whale bodies can be emerged or submerged (e.g. lobtailing, blowing, logging, etc.); and 5) occlusions and noise could occur due to clouds, aerosols, haze, sunglint, or water turbidity.

Objectives

In this work, we propose a robust and generalizable deep learning system for automatically counting whales from space. For this, we built a two-step CNN-based model, where the first step detects the presence of whales and the second step counts the number of whales in satellite images (Fig. 1). To overcome the above mentioned challenges, 1) we combined several open datasets to build an annotated training database of high quality orthoimages of whales and of objects that could be confused with whales, 2) we used data augmentation and transfer learning techniques to make the CNNs robust to image variability, 3) we assessed the effect of whale posture and location on model performance, and 4) we applied the model to free Google Earth coastal imagery in 10 whale watching hotspots as a proof of concept.

In brief, the first step CNN-based model was applied on a 71 m x 71 m sliding windows -twice the size of blue whales (30 m)- in 10 whale hotspots around the world and outputs the probability of having detected whales in each window (Fig. 1A). The second step CNN-based model analyzed only those windows with high probability of whale presence, localized each whale within a bounding box, and output the number of counted individuals (Fig. 1B). To facilitate its use and to support whale conservation, the CNN-based model was built using open-source software and can be used on free Google Earth images (subjected to terms of service).

Fig. 1. The proposed automatic whale-counting procedure with a two-step CNN-based model. **(A)** The first-step CNN scans the sample area (following the yellow line) to search for the presence of whales in each grid cell (white squares). Only grid cells where the first-step CNN gives high probability for whale presence (red square) are analyzed by **(B)** the second-step CNN, which finally locates and counts individuals (green bounding boxes).

Methods summary

Training, testing and validating datasets

Since there not exist accessible datasets of orthoimages for whales yet, we had to build two datasets for training the CNN-based models to respectively detect the presence of whales and count their number, and a third dataset for testing and validating the whole procedure. The three datasets were built by combining, preprocessing and labeling images from several sources: Google Earth (28), free Arkive (29), NOAA Photo Library (30), and NWPU-RESISC45 dataset (31). For the first step, the training dataset had 2,800 images of the following three object classes (700 images per class): 1) whales, 2) ships, and 3) “water + submerged rocks” (Data S1). For the second step, the same training dataset with 700 images containing whales was used, but each whale was annotated within a bounding box (the total number of bounding boxes was 945).

The dataset for testing and validating the whole procedure consists of RGB satellite images downloaded from Google Earth in 14,148 cells of 71 m x 71 m distributed worldwide. For ships, we selected 400 images from 100 seaports around the world (Fig. 2A, Data S2). For “water + submerged rocks” class, we selected 400 coastal images randomly around the world (Fig. 2A, Data S3). Finally, for whales (Table 1), we downloaded 13,348 cells (Data S4) of 71 m x 71 m from 10 areas that had very high-resolution images at zoom 18 and that are known for marine mammal diversity or whale watching. These areas have been highlighted either as global marine biodiversity hotspots (32), marine mammal hotspots (33), or irremplazable or priority conservation sites (11), and are included within or next to a marine protected area (34) (Table 1). Two authors visually inspected all the images to annotate each cell with the name of the corresponding class and with the number of whales. From the 13,348 cells in the 10 hotspots for whale watching, the authors’ visual photo-interpretation revealed whales only in 70 cells.

Table 1. Summary of the inputs and outputs of the whale presence detection model (step 1).

Location of the ten sites evaluated. Match between the known whale watching period from the literature and the acquisition date and season of the satellite images in Google Earth. Number of 71 m x 71 m grid cells where the model and the authors detected whale presence and absence. For each image source in Google Earth these metadata are provided: the satellite (GE-01: GeoEye-01; QB-02: QuickBird-2; SPOT-6; USGS: United States Geological Survey orthoimagery; WV-02: WorldView-2; and WV-03: WorldView-3), the pixel size at nadir in meters, and the sensor spectral resolution (M: Multispectral; P: Panchromatic).

Site IDs Site names (country)	Latitude Longitude WGS84	Whale watching period	Date of Google Earth image	Season of acquisition date	Image source Pixel size (m)	Cells with whales (photo- interpreted)	Cells with whales (model)	Cells without whales (model)
1. Hawaiian Islands (USA)	20.636602 156.462511	-Dec-Apr (41)	Apr 3, 2013 Jan 13, 2013	Breeding	USGS aerial 0.15 M	4	4	565
2. Baja California (Mexico)	26.769961 113.242382	-Feb (42)	Feb 20, 2017	Breeding	WV-03 0.31 P 1.24 M	7	6	2,974
3. Valdés Peninsula (Argentina)	-42.603384 64.810850	-May-Dec (43)	Sep 18, 2003	Breeding	QB-02 0.61 P 2.5 M	7	4	1,295
4. Witsand (South Africa)	-34.390203 20.879985	Jul-Oct (4) (4)	Aug 9, 2009	Breeding	GE-01 0.46 P 1.84 M	41	35	3,312
5. Memba (Mozambique)	-14.185282 40.691405	Jun-Jul (43) (43)	Jun 23, 2017	Breeding	SPOT-6 1.5 P 6.0 M	1	1	594
6. Coral Sea (Australia)	-24.622170 153.291559	Sep-Nov (44)	Sep 13, 2005	Breeding	QB-02 0.61 P 2.5 M	2	2	346
7. Enderby Island (New Zealand)	-50.501698 166.282294	Jul-Sep (45) (45)	Sep 2, 2012	Breeding	WV-02 0.46 P 1.84 M	8	4	790
8. Peruvian coast (Peru)	-14.253483 76.159243	-Jun-Sep (46)	Mar 9, 2016	Migrating/ Feeding	WV-03 0.3 P 1.24 M	0	0	1,307
9. Canary Islands (Spain)	28.139039 16.796631	-Aug-Nov (47)	Mar 10, 2017	Migrating	WV-02 0.46 P 1.84 M	0	0	1,045
10. Japanese coast (Japan)	41.947425 143.246413	Apr (48) (48)	Oct 5, 2014	Migrating	WV-02 0.46 P 1.84 M	0	0	1,120
Total						70	56	13,348

Metrics used in the performance assessment

To evaluate the performance of both CNN-based models, we used these metrics (35): positive predictive value, sensitivity, and F1-measure.

$$\text{Positive predictive value} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}}$$

$$\text{Sensitivity} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$

$$\text{F1 measure} = 2 \times \frac{\text{positive predictive value} \times \text{sensitivity}}{\text{positive predictive value} + \text{sensitivity}}$$

True positives correspond to images that were correctly classified or counted as whales by the models, false positives correspond to images that were classified or counted as whales by the models but actually corresponded to another class, and false negatives correspond to undetected images with whales. In simple terms, high positive predictive value means that the model returned substantially more actual whales than false ones, while high sensitivity means that the model returned most of the actual whales. F1-measure provides a balance between precision and sensitivity.

Step 1: Whale presence detection phase

When seen from space, whales are often confused with other object classes such as ships and wave foam around partially or entirely submerged rocks. To give the first step CNN-based model the capacity to distinguish between these objects, we addressed the problem as a three-class image classification task. The first model was built using the last version of GoogleNet Inception v3 CNN architecture (36), pretrained on the massive ImageNet database (around 1.28 million images, organized into 1,000 object categories). We only removed the two last learnable layers from the network and re-trained them with our dataset, fine-tuning the parameters. To assess whether whale posture, season, and location affected whale presence

detection in satellite images, we compared the F1-measure metric across different seasons and locations of the world, and across multiple active and resting swimming movements (37), i.e., peduncle, lobtailing, breaching, blowing, spyhopping, logging, and submerged (38).

Step 2: Whale counting phase

We built the second CNN-based model that counts whales by reformulating the problem into an object detection task. We used the detection model Faster R-CNN based on Inception-Resnet v2 CNN architecture (25, 39), pre-trained on the well known COCO (Common Objects in Context) detection dataset, which contains more than 200,000 images organized into 80 object categories (40).

A) Global testing input

B) Output from the whale presence detection step

C) Output from the whale counting step

Fig. 2. Study sites and illustration of the results from the whale detection and counting models. **(A)** Distribution of 10 marine mammal hotspots (details in Table 1) for whale watching. Red, blue, and yellow cells indicate whale presence, water + submerged rocks, and ships, respectively. **(B)** Detail of the 71 m x 71 m grid cells assessed in coastal areas worldwide. The first step CNN-based model detected the presence of whales in seven of the 10 candidate hotspots (in the three remaining hotspots, high resolution images were not available for the whale watching months). **(C)** Output from the whale counting (Step 2). Detail of the outputs from the second step CNN-based model that located and counted the number of whales (green bounding boxes) in the grid cells where the first-step CNN gave high probability for whale presence.

Results

Whale presence detection and step-1 model validation

The first step CNN-based model that detects the presence of whales reached an average F1-measure of 84% for whales, and of 96% and 97% for the background classes water + submerged rocks and ships, respectively (Table S2, Fig. 3). Only 21.12% of test grid cells containing whales were misclassified as water (18.30 %) or ships (2.82 %). A very small number of water + submerged rocks and ship images were misclassified as whales (1.00 % and 2.25 %, respectively; see Fig. 3).

Fig. 3. Visualization (Circos plot) of the confusion matrix between the photo-interpreted ground truth and the predictions made by the first step CNN-based model for detecting the presence of whales (in red), ships (yellow), and water + submerged rocks (blue). Only 13 and 1 whale images were classified as water + submerged rocks and as ships, respectively, while only 9 ship images and 4 water + submerged rocks images were classified as whales.

The first step CNN-based model confirmed the presence of whales in seven of the ten whale watching hotspots assessed (see Data S4, Table 1, Fig. 2). The acquisition date of the

satellite images available in Google Earth for these seven sites matched the known whale watching period from the literature (41-48) (Table 1). In the three sites where the model did not find whales (Peru, Canary Islands, and Japan), the acquisition date of the Google Earth image was not within the known whale watching period but during the migration season. In the Peruvian coast and in the Canary Islands the detection was particularly challenging since the images presented rough sea. However, the author's photo-interpretation confirmed the absence of whales in the images of Peru, Canary Islands, and Japan (Table 1).

Swimming movements affected the performance of the first step CNN-based model to detect the presence of whales (Fig. 4). Higher detectability (greater than 90% of true positives) was obtained for the following whale postures: blowing, breaching, lobtailing, peduncle, and logging. The lowest detectability occurred for submerged and spyhopping postures (33% and 60% of false negative, respectively; see Fig. 4A). Indeed, the lower performance of the step one model in the Argentinean and New Zealand sites (Table S1, Fig. S1) was due to the much greater frequency of these latter postures in the images (Data S5). Overall, greater number of whales were in the passive swimming movements of logging and submerged (60% of detected whales and 74% of photo-interpreted whales), while the lower number of whales were detected under active movements (Fig. 4A, Data S5).

Fig. 4. Impact of whale postures or swimming movements on the performance of the Step 1 CNN-based model for detecting the presence of whales. **(A)** The Circos plot shows the distribution of the false negatives (undetected whales, in red color) and true positives across whale postures. Whales under blowing, breaching, lobtailing, and peduncle postures were better detected than under logging, submerged, and spyhopping. **(B)** Example of images for each swimming movement from the detected hotspots (Fig. 2B) at the finest zoom.

Whale counting and step-2 model validation

The second step CNN-based model localized and counted whales in all the cells where the first step found whale presence, reaching $97 \pm 0.04\%$ of F1 measure (Fig. 2C, Table S1). Less than 2% of whales in these cells were not localized and counted by the second step model (one in South Africa). From a total number of 76 whales photo-interpreted

in this study across seven hotspots for whale watching around the world, the two-step CNN-based model automatically localized and counted 62 of them, which gives the model an overall global performance of $84 \pm 0.13\%$ of F1 measure (Table S1).

Discussion and conclusions

This study illustrates how global cetacean conservation could benefit from the operational application of deep learning on very high resolution satellite images. Using a two-step convolutional neural network model trained with a reduced dataset and applied on free Google Earth imagery, we managed to automatically detect and count 62 whales in seven hotspots for whale watching around the world, reaching an overall global F1-measure of 89% (F1 measure of 84% for presence detection and 97% for locating and counting). Our results show how the acquisition date of the satellite image and the swimming movement recorded in the image can influence whale presence detection and counting from space. This robust, transparent and automatic method can have direct and wide implications for whale conservation by assessing whale distributions, abundances, and movement behaviours globally from space.

Our satellite-based assessment can complement and be compared with other aerial, marine, and land observations. The coastal images of Google Earth at zoom 18 that we used correspond to a visual altitude of ~254 m, similar to the aerial surveys for grey whales, and up to ~4 km offshore the coast, the maximum distance for whale visual surveys from land (49). In whale assessments, such distances are good enough to get reliable estimates of instantaneous presence and relative population abundances (50). As new images become available, our method also enables dynamic updates at low cost, to assess seasonal and

interannual changes in population sizes, feeding and breeding areas, migratory routes, and distribution ranges around the world.

CNNs can outperform humans with good quality images

Several studies show that the performance of CNNs can be equal or even better than humans when the quality of the images is good, for instance, for skin cancer detection (51), mastering the game of Go (52), or generalizing past experiences to new situations (53). In general, the quality of the images determines the accuracy of the classification in CNNs (54), learning and performing better on higher resolution images (55). However, our results show how CNN-based methods trained on high-quality images (see methods section) can also reach good performance in classification and detection on medium-quality images, such as those available for free in Google Earth. In addition, the CNN-based models are robust (56) against the differences in spatial and spectral resolutions, and illumination angles across the different satellite sensors used in Google Earth (29).

Limitations of the CNN-based models

The use of free Google Earth imagery is convenient but it also has limitations since these are RGB images rather than multispectral, only available for few dates that may not be within the known whale presence period, are generally constrained to limited locations along coastal areas (up to ~4 km offshore), and are restricted for massive access. These last three limitations must be overcome together with the use of supercomputing for the worldwide “wall-to-wall” application of this method but do not impede its use for local assessments of whales around the world. Image spatial resolution can also limit the application of this method to detect cetaceans shorter than 5 m long (e.g. pilot whales, dolphins, etc.), which would require pixel sizes smaller than 1 m. For example, in our study, higher resolution

images tended to give a higher F1-measure (Table S3), though low contrast between whales and surrounding water tended to decreased performance (e.g. New Zealand) and high contrast to improve it (e.g. Mozambique) (Table S3).

Our results showed that the swimming movement and the image acquisition date can also bias the probability of detecting whales. The spatial pattern of whales under blowing, breaching, lobtailing, and peduncle postures showed better detectability than under logging and submerged, when whale bodies can be confused with submerged rocks and seafloor. However, the greater number of whales (both detected by the model and photo-interpreted) in our study were under passive (logging and submerged) instead of active swimming movements, and in images captured during the breeding season. Therefore, the best time to identify whales might be along the breeding season (Table 1), when whales spend more time in surface and in shallow waters (57). The effect of overlapped positions between females and calves on their detectability and counting should be further studied. In contrast, the most difficult time might be during migration and in the feeding season (Table 1), when whales are mainly in spyhopping, peduncle, lobtailing, and deeply submerged postures (58), and in areas with low contrast between water and whales, or under high sea surface roughness, sea glint, or bad atmospheric conditions (clouds or aerosols).

A future for conservation with CNNs applied to remote sensing data

The application of CNNs in remote sensing opens a world of possibilities for biodiversity science and conservation (37, 59). The great performance obtained by the CNN-based models trained on and applied to free high-resolution images opens the possibility to automatically process millions of satellite images around the world from whale hotspots, marine protected areas, whale sanctuaries, or migratory routes. Our procedure requires less

time and lower cost than the traditional acoustic surveys from ships or the visual surveys from planes and helicopters. The efficiency of remote sensing methods is particularly relevant to save time and money for long-term whale monitoring in remote places, or under difficult circumstances such as whales trapped inside sea ice in polar regions (60). The detection of whales using satellite images was already achieved using classical methods (19), but their portability to other regions or dates was strongly limited by the necessity of spectral normalization. However, our CNN-based model is easily transferable to any region or image with different characteristics in color, lighting and atmospheric conditions, background, or size and shape of the target objects, and it requires low human supervision, which speeds up the detection process (36).

Further research could increase the performance and variety of species identified by our CNN-model. For instance, the model could be improved by increasing the number of samples and variety of atmospheric and sea conditions in the training datasets, by building hierarchical training datasets with different swimming movements across different species (61), by using more spectral bands and temporal information (62), and by artificially increasing the spatial resolution of the images through rendering (63). In addition, as it is a fast and scalable method, it can even be transferred to very high spatial resolution images (<10 cm) captured by unmanned aerial vehicles (UAVs) for the automatic identification of specific individuals (64).

A global operationalization of our satellite-based model for whale detection and counting could greatly complement traditional methods (13-21) to assist whale conservation, to guide marine spatial planning (65), or to assess regional (11) and global (33) priorities for marine biodiversity protection against global change (66). In addition, since deep learning is

agnostic to the type of image data used, our method could be adapted to identify and quantify other marine species such as seals and sea lions (67), penguins (68), etc. To boost this process, free access to satellite data is key (69). The compromise with biodiversity conservation from corporations such as Google, Microsoft, Planet, Airbus, or DigitalGlobe (70) could be materialized through the systematic release of free high resolution aerial and satellite imagery at least from key sites for marine conservation. Even more, the acquisition of these images in pelagic environments does not directly compete with satellites commercial activity, which is usually focused on terrestrial and coastal areas. Having these images available would also make it possible to organize the development of a global database of orthoimages of cetaceans and many other marine vertebrates that could be used to improve the training of our whale detection and counting model or to develop similar models for other marine organisms. Images of the highest spatial resolution (such as WorldView-3 satellite images with a pixel size of 0.3 m) are particularly appropriate for this purpose. This way, satellite and CNN-based detectors of big marine organisms could serve to produce global characterizations of species populations and traits and of community composition as part of the initiative by the Group on Earth Observations - Biodiversity Observation Network (GEOBON) on satellite remote sensing essential biodiversity variables (71).

REFERENCES

1. Payne R, Rowntree V, Perkins JS (1990) Population size, trends and reproductive parameters of right whales (*Eubalaena australis*) off Peninsula Valdes, Argentina. *Rep. Int. Whal. Commn.* 12, 271-8.
2. Corkeron PJ (2004) Whale watching, iconography, and marine conservation. *Conserv. Biol.* 18, 847-849.
3. Parsons ECM, *et al.* (2015) Key research questions of global importance for cetacean conservation. *Endanger. Species Res.* 27, 113-118.
4. O'Connor S, Campbell R, Cortez H, Knowles T (2009) Whale Watching Worldwide: tourism numbers, expenditures and expanding economic benefits, a special report from the International Fund for Animal Welfare. *Yarmouth MA, USA, prepared by Economists at Large* 228.
5. Bowen WD (1997) Role of marine mammals in aquatic ecosystems. *Mar. Ecol. Prog. Ser.* 158, 267-274.
6. Smith LV, *et al.* (2013) Preliminary investigation into the stimulation of phytoplankton photophysiology and growth by whale faeces. *J. Exp. Mar. Biol. Ecol.* 446, 1-9.
7. Roman J, *et al.* (2014) Whales as marine ecosystem engineers. *Front. Ecol. Environ.* 12, 377-385.
8. Jenkins L, Romanzo C (1998) Makah whaling: aboriginal subsistence or a stepping stone to undermining the commercial whaling moratorium. *Colo. J. Int'l Envtl. L. & Pol'y.* 9, 71.
9. Jetz W, Freckleton RP (2015) Towards a general framework for predicting threat status of data-deficient species from phylogenetic, spatial and environmental information. *Philos. T. R. Soc. B.* 370, 20140016.

10. IUCN (2017) “IUCN red list of threatened species”. Ver. 2017.2. www.iucnredlist.org,
Downloaded on December.
11. Jenkins CN, Van Houtan KS (2016) Global and regional priorities for marine biodiversity protection. *Biol. Conser.* 204, 333-339.
12. Oleson EM, Wiggins SM, Hildebrand JA (2007) Temporal separation of blue whale call types on a southern California feeding ground. *Anim. Behav.* 74, 881-894.
13. Mellinger DK, Stafford KM, Moore SE, Dziak RP, Matsumoto H (2007) An overview of fixed passive acoustic observation methods for cetaceans. *Oceanography* 20, 36-45.
14. Barlow J, Taylor BL (2005) Estimates of sperm whale abundance in the northeastern temperate Pacific from a combined acoustic and visual survey. *Mar. Mammal Sci.* 21, 429-445.
15. Kennedy AS, *et al.* (2013) Local and migratory movements of humpback whales (*Megaptera novaeangliae*) satellite-tracked in the North Atlantic Ocean. *Can. J. Zool.* 92, 9-18.
16. Wade P, *et al.* (2006) Acoustic detection and satellite-tracking leads to discovery of rare concentration of endangered North Pacific right whales. *Biology Lett.* 2, 417-419.
17. Schweder T, Sadykova D, Rugh D, Koski W (2010) Population estimates from aerial photographic surveys of naturally and variably marked bowhead whales. *J. Agric. Biol. Envir. S.* 15, 1-19.
18. Hodgson A, Kelly N, Peel D (2013) Unmanned aerial vehicles (UAVs) for surveying marine fauna: a dugong case study. *PloS One* 8, e79556.
19. Fretwell PT, Staniland IJ, Forcada J (2014) Whales from space: counting southern right whales by satellite. *PLoS One* 9, e88655.
20. R. Abileah R (2002) Marine mammal census using space satellite images. *US Navy Journal of Underwater Acoustics* 52.

21. Aniceto AS, *et al.* (2018) Monitoring marine mammals using unmanned aerial vehicles: quantifying detection certainty. *Ecosphere* 9.
22. Taylor BL, Martínez M, Gerrodette T, Barlow J, Hrovat YN (2007) Lessons from monitoring trends in abundance of marine mammals. *Mar. Mammal Sci.* 23, 157-175.
23. Kerr JT, Ostrovsky M (2003) From space to species: ecological applications for remote sensing. *Trends Ecol. Evol.* 18, 299-305.
24. Rose RA, *et al.* (2015) Ten ways remote sensing can contribute to conservation. *Conserv. Biol.* 29, 350-359.
25. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. *Adv. Neur. In.* 1097-1105.
26. Ren S, He K, Girshick R, Sun J (2017) Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE T. Pattern Anal.* 39, 1137-1149.
27. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521, 436–444.
28. Yu L, Gong P (2012) Google Earth as a virtual globe tool for Earth science applications at the global scale: progress and perspectives. *Int. J. Remote Sens.* 33, 3966-3986.
29. Wildscreen Arkive (2018) Discover the world's most endangered species.
www.arkive.org, Downloaded on January.
30. NOAA (2018) Photo Library. www.photolib.noaa.gov, Downloaded on January.
31. Cheng G, Han J, Lu X (2017) Remote sensing image scene classification: benchmark and state of the art. *P. IEEE* 105, 1865-1883.
32. Ramírez F, Afán U, Davis LS, Chiaradia A (2017) Climate impacts on global hot spots of marine biodiversity. *Sci. Adv.* 3, e1601198.
33. Pompa S, Ehrlich PR, Ceballos G (2011) Global distribution and conservation of marine mammals. *Proc. Natl. Acad. Sci. U.S.A.* 108, 13600-13605.

34. UNEP-WCMC, IUCN (2018) Protected Planet: The World Database on Protected Areas (WDPA) [On-line], July 2018, Cambridge, UK: UNEP-WCMC. Available at www.protectedplanet.net.
35. Guirado E, Tabik S, Alcaraz-Segura D, Cabello J, Herrera F (2017) Deep-learning versus OBIA for scattered shrub detection with Google Earth images: *Ziziphus lotus* as case study. *Remote Sens-basel*. 9, 1220.
36. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2016) Rethinking the inception architecture for computer vision. *Proc. CVPR IEEE* 2818-2826.
37. Lyamin OI, Manger PR, Mukhametov LM, Siegel JM, Shpak OV (2000) Rest and activity states in a gray whale. *J. Sleep Res.* 9, 261-267.
38. Kavanagh AS, *et al.* (2017) Evidence for the functions of surface-active behaviors in humpback whales (*Megaptera novaeangliae*). *Mar. Mammal Sci.* 33, 313-334.
39. Szegedy C, Ioffe S, Vanhoucke V (2016) Inception-v4, inception-resnet and the impact of residual connections on learning. Preprint at <https://arxiv.org/abs/1602.07261>.
40. Lin TY, *et al.* (2014) Microsoft coco: Common objects in context. *Comput. Vis. ECCV* 740-755.
41. Baird RW, Burkhart SM (2000) Bias and variability in distance estimation on the water: implications for the management of whale watching. In *IWC Meeting Document SC/52/WW1*.
42. Seger KD, Thode AM, Swartz SL, Urban RJ (2015) The ambient acoustic environment in Laguna San Ignacio, Baja California Sur, Mexico. *J. Acoust. Soc. Am.* 138, 3397-3410.
43. Banks A, *et al.* (2011) Recent sightings of southern right whales in Mozambique. Paper SC/S11/RW17 presented to the *IWC Southern Right Whale Assessment Workshop*, Buenos Aires, Argentina. 21 pp.

44. Stamation KA, Croft DB, Shaughnessy PD, Waples KA, Briggs SV (2010) Behavioral responses of humpback whales (*Megaptera novaeangliae*) to whale-watching vessels on the southeastern coast of Australia. *Mar. Mammal Sci.* 26, 98-122.
45. Stewart R, Todd B (2001) A note on observations of southern right whales at Campbell Island, New Zealand. *J. Cetacean Res. Manage.* 2, 117-120.
46. Felix F, Haase B (2005) Distribution of humpback whales along the coast of Ecuador and management implications. *J. Cetacean Res. Manage.* 7, 21.
47. Carrillo M, Pérez-Vallazza C, Álvarez-Vázquez R (2010) Cetacean diversity and distribution off Tenerife (Canary Islands). *Mar. Bio. Rec.* 3, 1-9.
48. Darling JD, Cerchio S (1993) Movement of a humpback whale (*Megaptera novaeangliae*) between Japan and Hawaii. *Mar. Mammal Sci.*, 9, 84-88.
49. Shelden KE, Laake IL (2002) Comparison of the offshore distribution of southbound migrating gray whales from aerial survey data collected off Granite Canyon, California, 1979-96. *J. Cetacean Res. Manag.* 4, 53-56.
50. Rowat D, Gore M, Meekan MG, Lawler IR, Bradshaw CJ (2009) Aerial survey as a tool to estimate whale shark abundance trends. *J. Exp. Mar. Bio. Ecol.* 368, 1-8.
51. Esteva A, *et al.* (2017) Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542, 115.
52. Silver D, *et al.* (2016) Mastering the game of Go with deep neural networks and tree search. *Nature* 529, 484–489.
53. Mnih V, *et al.* (2015) Human-level control through deep reinforcement learning. *Nature* 518, 529–533.
54. Dodge S, Lina K (2016) Understanding how image quality affects deep neural networks. *Quality of Multimedia Experience (QoMEX), 2016 Eighth International Conference on.* IEEE.

55. Kim J, Lee JK, Lee KM (2016) Accurate image super-resolution using very deep convolutional networks. *Proc. CVPR IEEE*, pp. 1646-1654.
56. Tabik S, *et al.* (2018) A snapshot of image pre-processing for convolutional neural networks: case study of MNIST. *Int. J. Comput. Int. Sys.* 10, 555–568.
57. Dawbin WH (1966) The seasonal migratory cycle of humpback whales in Whales, dolphins and porpoises. K. S. Norris, ed. (U. C. Press, Berkeley, CA, 1966), pp 145–171.
58. Corkeron PJ (1995) Humpback whales (*Megaptera novaeangliae*) in Hervey Bay, Queensland: behaviour and responses to whale-watching vessels. *Can. J. Zool.* 73, 1290-1299.
59. Su JH, Piao YC, Luo Z, Yan BP (2018) Modeling Habitat Suitability of Migratory Birds from Remote Sensing Images Using Convolutional Neural Networks. *Animals*, 8, 66.
60. Williams R, *et al.* (2014) Counting whales in a challenging, changing environment. *Sci. Rep-UK* 4, 4170.
61. Yan Z, *et al.* (2015) HD-CNN: hierarchical deep convolutional neural networks for large scale visual recognition. *Proc. Int. Conf. on Computer Vision*, pp. 2740-2748.
62. Basaeed E, Bhaskar H, Al-Mualla M, (2015) CNN-based multi-band fused boundary detection for remotely sensed images. In Proceedings of the International Conference on Imaging for Crime Prevention and Detection, London, UK, 15–17 July 2015.
63. Fang L, Au OC, Tang K, (2012) "Increasing image resolution on portable displays by subpixel rendering—A systematic overview", *APSIPA Trans. Signal Inf. Process.* 1, e1.
64. Apprill A, *et al.* (2017) Extensive Core Microbiome in Drone-Captured Whale Blow Supports a Framework for Health Monitoring. *MSystems* 2, e00119-17.
65. Augé A, *et al.* (2018) Framework for mapping key areas for marine megafauna to inform Marine Spatial Planning: The Falkland Islands case study. *Mar. Policy* 92, 61-72.

66. Hays GC, *et al.* (2016) Key questions in marine megafauna movement ecology. *Trends Ecol. Evol.* 31, 463-475.
67. Moxley JH, *et al.* (2017) Google haul out: Earth observation images and digital aerial surveys in coastal wildlife management and abundance estimation. *BioScience* 67, 760-768.
68. Lynch HJ, White R, Black AD, Naveen R (2012) Detection, differentiation, and abundance estimation of penguin species by high-resolution satellite images. *Polar Biol.* 35, 963-968.
69. Buchanan GM, *et al.* (2018) Free satellite data key to conservation. *Science*, 361, 139-140.
70. Popkin G (2018) Technology and satellite companies open up a world of data. *Nature* 557, 745-747.
71. Pettorelli N, Owen HJF, Duncan C (2016) How do we want Satellite Remote Sensing to support biodiversity conservation globally? *Methods Ecol. Evol.* 7, 656-665.
72. Abadi M, *et al.* (2016) TensorFlow: large-scale machine learning on heterogeneous systems. Preprint at <https://arxiv.org/abs/1603.04467>.
73. Huang J, *et al.* (2017) Speed/accuracy trade-offs for modern convolutional object detectors. *Proc. CVPR IEEE*, 4.

Acknowledgments

Funding: S.T. was supported by the Ramón y Cajal Programme of the Spanish government (RYC-2015-18136). S.T., E.G., and F.H. were supported by the Spanish Ministry of Science under the project TIN2017-89517-P. D. A-S. received support from European LIFE Project ADAPTAMED LIFE14 CCA/ES/000612, and from ERDF and Andalusian Government under the project GLOCHARID. D.A-S. received support from NASA Work Programme on Group on Earth Observations - Biodiversity Observation Network (GEOBON) under grant 80NSSC18K0446, from project ECOPOTENTIAL, funded by European Union Horizon 2020 Research and Innovation Programme under grant agreement No. 641762, and from the Spanish Ministry of Science under project CGL2014-61610-EXP and grant JC2015-00316. M.R. received support from International mobility grant for prestigious researchers by (CEIMAR) International Campus of Excellence of the Sea. This work was partially funded by the project BigDaP-TOOLS - Ayudas Fundación BBVA a Equipos de Investigación Científica 2016.

Author contributions: E.G. and S.T. designed the experiments, collected the data, implemented the code, ran the models, analyzed the results, and wrote the paper. D.A-S., and M.L.R., provided guidance, contributed to the design of the analyses, analyzed the results, and wrote the paper. F.H. provided guidance, contributed to the design of the study, and wrote the paper.

Competing interests: The authors declare that no competing interests exist in relation to this manuscript.

Data and material availability statement: The test and training datasets that support our findings are available from Github archive (<https://github.com/EGuirado/CNN-Whales-from-Space>). Restrictions apply to the availability of the images of the training and validation data, which were used with permission for the current study, but are not publicly available. Some data may be available from the authors upon reasonable request and under written permission from Google Earth, Arkive, or NOAA Photo Library when applicable. To allow reproducibility, in the supplementary information, we provide the metadata of all the images in the training and testing datasets.

The supplementary material listed below is provided under request via email: eguirado@ugr.es ; siham@ugr.es

- Tables S1 to S3
- Fig. S1
- Data S1 to S5