






ORIGINAL RESEARCH

Using very-high-resolution satellite imagery and deep learning to detect and count African elephants in heterogeneous landscapes

Isla Duporge¹ , Olga Isupova² , Steven Reece³ , David W. Macdonald¹  & Tiejun Wang⁴ ¹Wildlife Conservation Research Unit, Department of Zoology, University of Oxford, Recanati-Kaplan Centre, Tubney, UK²Department of Computer Science, University of Bath, Bath, UK³Department of Engineering Science, University of Oxford, Oxford, UK⁴Faculty of Geo-information Science and Earth Observation (ITC), University of Twente, Enschede, The Netherlands**Keywords**

Machine Learning, Convolutional Neural Network, Aerial Survey, Wildlife Census, Endangered Species, Conservation, Anthropocene, Object Detection

CorrespondenceIsla Duporge, Wildlife Conservation Research Unit, Department of Zoology, University of Oxford, Recanati-Kaplan Centre, Tubney, UK.
E-mail: Isla.duporge@zoo.ox.ac.uk

Editor: Nathalie Pettorelli

Associate Editor: Graeme Buchanan

Received: 11 September 2020; Revised: 31 October 2020; Accepted: 4 December 2020

Isla Duporge and Olga Isupova authors are equal contribution

doi: 10.1002/rse2.195

Remote Sensing in Ecology and Conservation 2021;7 (3):369–381**Abstract**

Satellites allow large-scale surveys to be conducted in short time periods with repeat surveys possible at intervals of <24 h. Very-high-resolution satellite imagery has been successfully used to detect and count a number of wildlife species in open, homogeneous landscapes and seascapes where target animals have a strong contrast with their environment. However, no research to date has detected animals in complex heterogeneous environments or detected elephants from space using very-high-resolution satellite imagery and deep learning. In this study, we apply a Convolution Neural Network (CNN) model to automatically detect and count African elephants in a woodland savanna ecosystem in South Africa. We use WorldView-3 and 4 satellite data –the highest resolution satellite imagery commercially available. We train and test the model on 11 images from 2014 to 2019. We compare the performance accuracy of the CNN against human accuracy. Additionally, we apply the model on a coarser resolution satellite image (GeoEye-1) captured in Kenya, without any additional training data, to test if the algorithm can generalize to an elephant population outside of the training area. Our results show that the CNN performs with high accuracy, comparable to human detection capabilities. The detection accuracy (i.e., F2 score) of the CNN models was 0.78 in heterogeneous areas and 0.73 in homogeneous areas. This compares with the detection accuracy of the human labels with an averaged F2 score 0.77 in heterogeneous areas and 0.80 in homogeneous areas. The CNN model can generalize to detect elephants in a different geographical location and from a lower resolution satellite. Our study demonstrates the feasibility of applying state-of-the-art satellite remote sensing and deep learning technologies for detecting and counting African elephants in heterogeneous landscapes. The study showcases the feasibility of using high resolution satellite imagery as a promising new wildlife surveying technique. Through creation of a customized training dataset and application of a Convolutional Neural Network, we have automated the detection of elephants in satellite imagery with accuracy as high as human detection capabilities. The success of the model to detect elephants outside of the training data site demonstrates the generalizability of the technique.

Introduction

The planet is in the geological era of the Anthropocene, during which human activity is the driving force of

change. Many wildlife species are under threat across their geographical range as we are currently undergoing the sixth-mass extinction (Barnosky et al., 2014; Cardinale, 2012; Skogen et al., 2018). Reliable, accurate, and up-to-

date data on wildlife numbers is essential to monitor population fluctuations and identify causes of decline. Various methods are used for conducting population counts, including, but not limited to line transect surveys (Emmanuel et al., 2017), dung and track counts (Barnes et al., 1997), bio-acoustic monitoring (Shiu et al., 2014), camera trap grids (Smit et al., 2017) and aerial surveys (Schlossberg et al., 2016), among others.

Satellite remote sensing has recently emerged as a new viable monitoring technique for detecting wildlife. It has been used to successfully identify and count several wildlife species in open, homogeneous landscapes and seascapes. The benefits of this monitoring technique are numerous; large spatial extents can be covered in short time periods making repeat surveys and reassessments possible at short intervals. For example, the satellite used in this study, Worldview-3 has an average revisit time of less than one day. It can collect up to 680,000 square kilometres every 24 h. Satellite images are captured over large areas in one shot so issues with double counting and miscounts are largely eliminated. Satellite remote sensing is unobtrusive as it requires no human presence, eliminating the risk of disturbing the species being surveyed. Disturbance remains a key concern in other surveying techniques (Mulero-Pazmany et al., 2017). Image acquisition is automated and less logistically complicated compared with traditional aerial surveys (Stapleton et al., 2014) and setting up camera trap grids or audio loggers. Censuses can be carried out without concern for human safety providing an ability to survey previously inaccessible areas. For example, in the case of the Emperor penguin, new colony locations were detected in a pan-continental survey of the Antarctic coast (Abileah, 2002; Smit et al., 2017). Additionally, cross border areas can be surveyed without requiring multiple national civil aviation permissions.

Detecting wild animals in satellite imagery is influenced by body size, background complexity and contrast between species and surrounding habitat. Seascapes provide a uniform high contrast background context against which whales have been identified in known breeding, calving and feeding grounds (Abileah, 2002; Cubaynes et al., 2018; Fretwell et al., 2014) and flamingos have been identified on a lake (Sasamal et al., 2008). Spectrally uniform rocky outcrops and ice have been used to identify several marine species, including Emperor and Adelie penguins (Barber-Meyer et al., 2007; Fretwell et al., 2012; Fretwell & Trathan, 2009; LaRue et al., 2014; Lynch & LaRue, 2014), Elephant and Weddell seals (LaRue et al., 2011; McMahon et al., 2014), Masked Boobies (Hughes et al., 2011) and Albatross (Fretwell et al., 2017). Several Arctic species have been identified against snow using shadow and body contrast for detection (e.g. Polar bears

LaRue & Stapleton, 2018; LaRue et al., 2017; LaRue et al., 2015; Stapleton et al., 2014) and muskoxen (LaRue et al., 2017)). On the African continent only two studies have detected mammals (wildebeest and zebra) using satellite imagery in open savannah (Xue et al., 2017; Yang et al., 2014), both in homogeneous monochrome environments. No study has yet, to the best of our knowledge, detected species in complex heterogeneous landscapes from space.

Various methods have been used to detect species in satellite imagery. The majority of studies have manually counted species in imagery using several observers for cross-validation. However, manual methods are unfeasible if large areas are surveyed, as counting is labour and time intensive and counts tend to be error-prone (Hollings et al., 2018). Several studies have relied on environmental proxies and indirect ecological signs of animal presence e.g. burrows (Löffler & Margules, 1980), mounds (Velasco, 2009), changes in vegetation from nest signatures (Hughes et al., 2011) and faecal stains in the case of penguins (Barber-Meyer et al., 2007; Fretwell et al., 2012; Fretwell & Trathan, 2009; LaRue et al., 2014; Lynch & LaRue, 2014; Lynch et al., 2012). Image differencing is a technique where satellite images are captured in the same location at different times. This technique is used for environmental change detection (Lu et al., 2010) e.g. deforestation and land use change (Kiage et al., 2007; Kusimi, 2008; Meng & Meentemeyer, 2011), identification of fire (Carvalho Júnior et al., 2015; Meng & Meentemeyer, 2011), droughts (Buma & Lee, 2019; Rulinda et al., 2010) or floods (Oliveira et al., 2019; Thito et al., 2016). Three studies used short-time image differencing to detect polar bears from space (LaRue & Stapleton, 2018; LaRue et al., 2015; Stapleton et al., 2014). Image differencing is possible in cases where multi-temporal imagery is available, and species can be differentiated from static objects. e.g. rocks. Images can be acquired via targeted satellite tasking on specific days; however, this is more costly than using archive imagery. Cloud cover, environmental variability and changing sea states can impede ground visibility which is problematic when image differencing and tasking is used.

Several studies have applied a form of supervised or semi-supervised classification approaches to detect species in satellite imagery. One form of image segmentation using semi-supervised classification is thresholding method. Pixel values are classified relative to a set of threshold values that distinguish species from background. Thresholding method does not make use of geometric information but rather relies on spectral signatures (pixel value combinations). Thresholding method is reliant on the human classifier to set accurate thresholds which is helped by familiarity with the species and environment (Xue et al., 2017). This technique is effective in

homogeneous environments where species have strong spectral separability from background context. However, in cases where pixel values of species and background context are similar it is difficult to draw accurate distinctions.

The introduction of Convolutional Neural Networks (CNN) in machine learning has revolutionized the field of computer vision since 2012 (Krizhevsky et al., 2012). Machine learning has become a new essential tool used by ecologists to detect wildlife in imagery e.g. camera trap images, aerial survey images and unmanned aerial vehicle (UAV) images (Bruijning et al., 2018; Chabot & Bird, 2015; Ferreira et al., 2020; Petersen et al., 2019; Torney et al., 2019; Weinstein, 2018). However, automated detection of wildlife from satellite imagery is still in its infancy. To the best of our knowledge only three species have been detected in satellite imagery using deep learning: albatross (Bowler et al., 2019), whales (Borowicz et al., 2019; Guirado et al., 2019) and pack-ice seals (Gonçalves et al., 2020). Object detection applications are now easier to develop than ever before. High-performance off-the-shelf solutions have made machine learning solutions accessible to non-specialists. These techniques can now leverage massive image datasets e.g. ImageNet (>14 million images across 20,000 classes) obtaining significant performance improvements compared to previous methods based on manually engineered features (Russakovsky et al., 2015). A Convolutional Neural Network (CNN) is a deep learning artificial neural network architecture that has been extensively used for object detection and recognition in recent years. The 'deep' stems from the use of multiple layers in the network. In this study, we test whether it is possible to detect the world's largest terrestrial mammal – the African elephant – using deep learning via a CNN.

The population of African elephants (*Loxodonta africana*) has plummeted over the last century due to poaching, retaliatory killing from crop raiding and habitat fragmentation (Gara, 2016; Poulsen et al., 2017; Sibanda & Murwira, 2012). To ensure conservation is achieved accurate monitoring is vital. Inaccurate counts can result in misallocation of scarce conservation resources and misunderstanding population trends. Existing techniques are prone to miscounting. The most common survey technique for elephant populations in savannah environments is aerial counts from manned aircraft (Schlossberg et al., 2016). Aerial surveys are conducted either as total counts – flying closely spaced transects, or sample counts, covering 5–20% of an area and extrapolating to a larger area. Several studies have shown that observers on aerial surveys often miscount due to fatigue and visibility issues resulting in over-estimates (Caughley et al., 1976; Jachmann, 2002; Koneff et al., 2008). Aerial surveys can be costly, logistically challenging in terms of finding suitable runways and refuelling stops and time consuming in terms

of getting appropriate permissions. This is particularly relevant in cross-border areas where multiple national permissions are required. Remotely sensing elephants using satellite imagery and automating detection via deep learning may provide a novel avenue for surveying while mitigating several of the challenges outlined above.

In this study, we investigate the feasibility of using very-high-resolution satellite imagery to detect wildlife species in heterogeneous environments with deep learning. To test this, we use a population in Addo Elephant National Park, South Africa where herds move between open savannah habitat and closed heterogeneous woodland and thicket.

Materials and Methods

Study Site

Addo Elephant National Park in South Africa was chosen as the study site. It provides a spectrally complex heterogeneous background with a high concentration of elephants. The park is the third largest park in South Africa at 1640 km². Different areas of the park have been sectioned off for conservation purposes – elephants were identified in the Main Camp section of the park surrounding Hapoor Dam (Figure 1). The Main Camp is a combination of dense shrubland and low forest e.g. pork-bush (*Portulacaria afra*), White milkwood (*Sideroxylon inerme*), Cape leadwort (*Plumbago auriculata*) and open grassland (Kakembo et al., 2015; Tambling et al., 2012). Over six hundred elephants move between these habitats (Du Toit & O'Connor, 2014; Wilgen et al., 2016). Elephants cover themselves in mud to cool down and adopt a range of postures when foraging, playing, sleeping (Soltis et al., 2016; Wickler & Seibt, 1997) so their shape and colour is continually changing. The background environment is also changing as they move between open savannah and woodland and take shelter under trees in the mid-day sun.

The park has rainfall year-round (Fullman et al., 2017) and four seasons can be broadly delineated as early wet season (Oct–Dec), late wet season (Jan–March), early dry (Apr–Jun) and late dry season (July–Sept) (Du Toit & O'Connor, 2014; Wilgen et al., 2016). To ensure a representative and diverse sample of elephants in the park we include training and test labels from images captured in different seasons and years (Table 1).

Dataset generation and satellite image pre-processing

WorldView-3 and WorldView-4 are the highest resolution satellite imagery commercially available. They

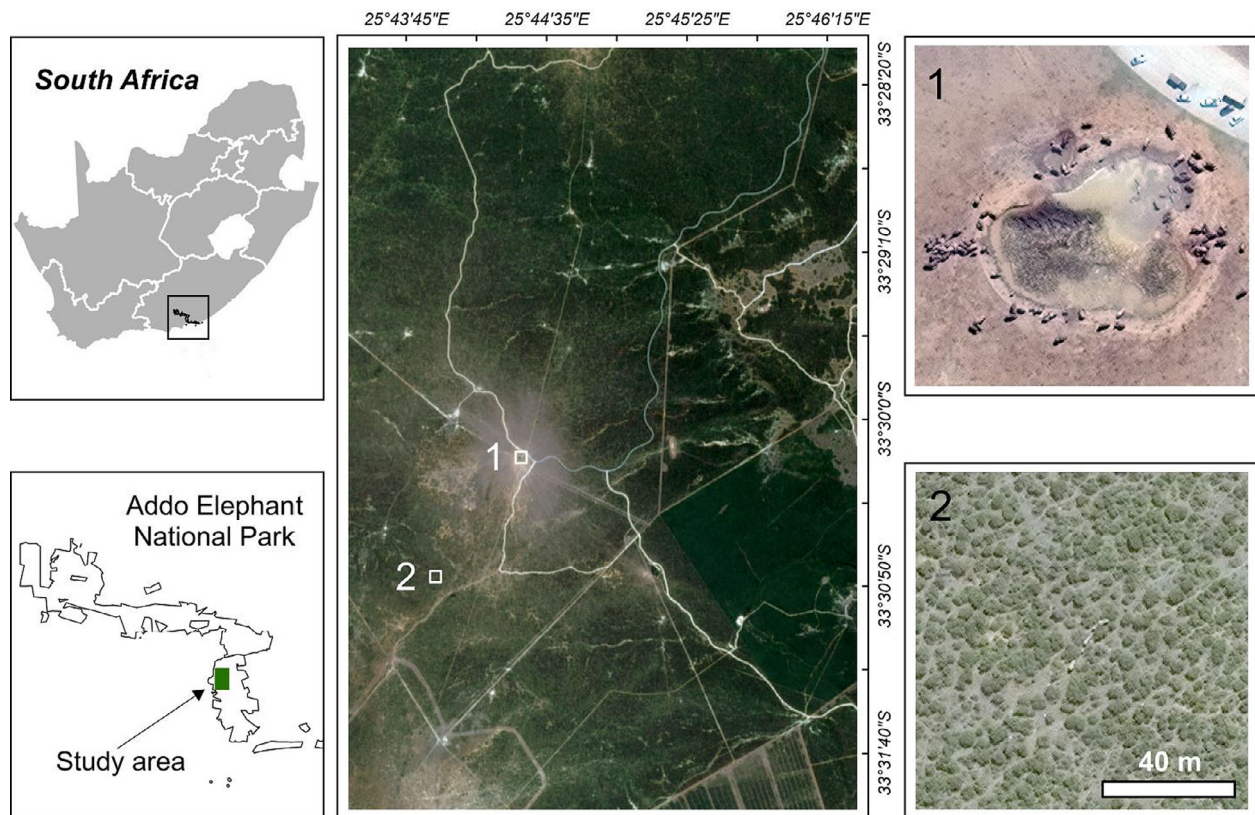


Figure 1. Location of the study area in -Addo Elephant National Park, South Africa. Two example WorldView-3 images showing 1) Elephants in open homogeneous area around Hapoor Dam, 2) Elephants in heterogenous woodland and thicket area. Satellite image (c) 2020 Maxar Technologies.

Table 1. List of satellite images used in the training and test dataset.

Date of acquisition	Satellite	Elephant labels in training dataset	Elephant labels in validation dataset	Elephant labels in test dataset
01_12_2014	WV3	197	52	10
29_01_2016	WV3	178	9	11
10_02_2016	WV3	259	19	32
03_04_2017	WV3	26	19	5
22_11_2017	WV4	10	0	0
11_01_2018	WV4	117	0	23
27_03_2018	WV4	236	16	24
08_10_2018	WV4	119	1	59
20_01_2019	WV3	22	0	0
11_08_2009	GeoEye-1	0	0	32

provide imagery at 31 cm resolution - WorldView-4 is now out of action but two years of archive imagery is available. The image archive for all WorldView 3 & 4 satellite images from Maxar Technologies (formerly DigitalGlobe) was searched via the Secure Watch Platform

[<https://www.digitalglobe.com/products/securewatch>]. We restricted the search to images that contain less than 20% cloud cover and acquired less than 25% off-nadir (degree off centre of image captured). We selected eleven images that met these specifications between 2014 and 2019. The satellite is on a sun synchronous orbital path, so each satellite image of the study area is captured between 10.10 and 10.45 am local time ensuring regular illumination conditions. The bright morning light improves image clarity as elephants gather at water holes in the morning which makes them easy to identify (Figure 1).

Each image was downloaded in two formats: orthorectified images in natural colour and orthorectified panchromatic image. We processed the images using a pan-sharpening algorithm from ERDAS IMAGINE software package (ERDAS, Inc., Atlanta, GA, USA) Pan-sharpening is an automatic image fusion process that uses the multi-spectral bands red (620–750 nm), green (495–570 nm), blue (450–495 nm) at 1.24 m resolution and the higher resolution panchromatic band at 31 cm to produce a high-resolution multispectral image. We tested several pan-sharpening algorithms using visual inspection method- the

Gram-Schmidt pan-sharpening algorithm provided the cleanest visual result in terms of spectral and spatial fidelity and was applied to all images. This is consistent with prior quantitative assessment of pan-sharpening algorithms that found Gram-Schmidt provides the highest spectral and spatial fidelity for identification of wildlife (Witharana et al., 2016). The satellite images were converted so that pixel values were in the range of [0,255] and the images were sliced into 600x600 pixel sub images to make them compatible with the deep learning software.

Labelling training data in satellite images

The images were visually scanned for elephants before sub-setting into smaller areas where we identified congregations of elephants. In total, 1125 elephants were identified in the training image dataset. To ensure training labels are representative of elephants at different times, images were selected for different seasons and years in both closed i.e. dense shrubland and forest, and open areas of the park i.e. grassland and bare land. Images were labelled by defining bounding boxes around each individual elephant using the graphical image annotation tool Labelling [https://github.com/tzutalin/labelImg] (Tzutalin, 2015) shown in Figure 2.

The baseline we deem as the true number of elephants is a labelled dataset doubled screened by two annotators – an Ecologist and Machine Learning Scientist. Any ambiguous labels that were not identified by both annotators were removed. We use the method of comparing the accuracy of detections from human volunteer annotators and CNN performance against this baseline control count (Ginosar et al., 2014; Torney et al., 2019). The same test images used to evaluate CNN performance were sent to 51 human volunteer annotators. The images were labelled by the volunteers using the VGG Annotation Tool [http://www.robots.ox.ac.uk/~vgg/software/via/] (Dutta & Zisserman, 2019). Volunteer annotators represent a cross-section of machine learning scientists, biologists, general public and park rangers who work with elephants in Southern Africa. The labellers vary in terms of computer literacy and familiarity with elephant behaviour and habitat preference. Each participant was provided with a detailed training sheet and an example of how elephants look in satellite images prior to labelling. The experiment involving human participants was approved by the University of Oxford CUREC Ethical Committee [R64699].

Training and validating the Convolutional Neural Network model

A CNN is a feed-forward neural network designed to process large-scale images by considering their local and

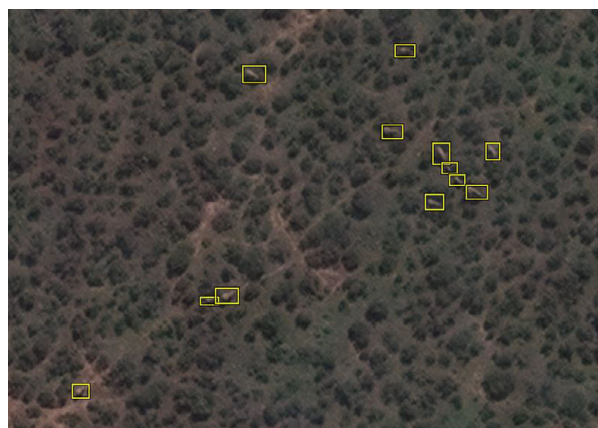


Figure 2. Example of elephant labels in a heterogeneous area, Addo Elephant National Park, South Africa. Satellite image (c) 2020 Maxar Technologies.

global characteristics (LeCun et al., 2015). A neural network is typically comprised of multiple layers connected by a set of learnable weights and biases (Romero et al., 2013). Convolutional layers represent a set of filters, each able to identify a particular feature in the image. These filters are fed small image patches whilst they scan across the image and generate feature maps for analysis by the next layer. The CNN comprises an alternating sequence of convolutional and pooling layers. Pooling layers are used to reduce the dimensionality of the feature maps to improve computational efficiency. Nonlinear activations are stacked between convolutional layers to enrich their representational power (Strigl et al., 2010). The last layer of the network is fully connected and performs classification (Schmidhuber, 2015). Convolutional neural networks have become a key tool in image classification. They are now comparable to human performance in a number of challenging image classification tasks e.g. face verification, various medical imaging tasks (Gulshan et al., 2016; Olczak et al., 2017).

We used the TensorFlow Object Detection API [https://github.com/tensorflow/models/tree/master/research/object_detection] to build our model (Huang et al., 2017). This API provides implementations of different deep learning object detection algorithms. In a preliminary assessment of the models available, we selected the model referred to as *faster_rcnn_inception_resnet_v2_atrous_coco* as it provided the best result and it was used for all the experiments presented. This model is a Faster Region CNN (RCNN) model – after layers that are used to extract features there is a subbranch to propose regions that may contain objects and a subbranch that predicts the final object class and bounding box for each of these regions (Ren et al., 2017). The model we used has an Inception

ResNet (Szegedy et al., 2017) backbone – this is the underlying CNN that is used for feature extraction. We used the model pretrained on the Common Objects in Context (COCO) dataset for object detection [https://cocodataset.org/] (Lin, 2014). We used default values for hyperparameters of this model from the API.

Training a CNN requires images to be split into training, validation and test sets. In total, 188 sub images from nine different satellite images were used for training. These training images contain 1270 elephant labels of which 1125 are unique elephants. There is an overlap of 50 pixels between sub images, when elephants appear at the edge of one sub image the overlap ensures they appear in whole on the neighbouring sub image. Twelve sub images containing 116 elephant labels were left out as a validation dataset. The validation dataset is used to tune the hyperparameters, to define the confidence threshold (above which predictions from the model are counted as detections) and to identify the optimal length of CNN training (see Figure 3).

Test dataset

The test dataset used to test the CNN against human annotator performance contains 164 elephants across seven different satellite images. These images do not overlap with any of the training or validation subimages. The test subimages cover both heterogeneous and homogeneous areas of the park from different seasons and years (see Table 1).

In addition, an image from the Maasai Mara in Kenya was used to test the generalizability of the CNN for broader image conditions - no additional training data were included for this test. The image comes from Geoeye-1 a lower resolution satellite (41 cm) compared

to the images used for training which come from World-view 3 & 4 (31 cm). The GeoEye-1 image was captured at about 10:30 am local time on 11th August 2009. This image allowed us to test the generalizability of our algorithm to a different environment and satellite.

Accuracy assessment

We compare the accuracy of detections from the human volunteer annotators and CNN against our count which we deem as the baseline i.e. the true number of elephants in the images. To calculate the deviation from this baseline we generate an F2 score.

Object detection performance is usually evaluated by *precision* and *recall*:

$$\text{Precision} = \frac{TP}{TP + FP} \text{ and } \text{Recall} = \frac{TP}{TP + FN}$$

where TP stands for true positives (correctly predicted elephants), FP stands for false positives (predicted elephants that are not actually elephants, also called false detections), FN – false negatives (elephants that are not detected by the model, also called missed detections).

The CNN gives an output in the form of bounding boxes, the same format we use for the training labels. We count any intersection between predicted and true bounding boxes as a true positive (i.e. the intersection over union threshold used to determine correct predictions was set to 0). Human volunteer annotators provide point detections for elephants. If these point detections were inside true bounding boxes, they were counted as true positives.

In precision and recall both types of errors – false positives and false negatives – are weighted equally. However, as it is more time consuming for a human annotator to check an entire image for missing elephants (false negatives) as compared with reviewing detected elephants and eliminating false positives we decided to use an F2 (F_β with $\beta = 2$) score. The F2 combines precision and recall in such a way that more emphasis is put on false negatives (Dascalu & David, 2019; Hordiiuk et al., 2019):

$$F_\beta = (1 + \beta^2) \frac{\text{Precision} * \text{Recall}}{\beta^2 * \text{Precision} + \text{Recall}} = \frac{(1 + \beta^2) TP}{(1 + \beta^2) TP + \beta^2 FN + FP},$$

which for $\beta = 2$ is equivalent to

$$F_2 = 5 \frac{\text{Precision} * \text{Recall}}{4 * \text{Precision} + \text{Recall}} = \frac{5TP}{5TP + 4FN + FP}$$

Performance of object detection algorithms are often measured by *average precision* (Huang et al., 2017) i.e. the area under a precision-recall curve that is obtained by varying the threshold of the confidence score. This

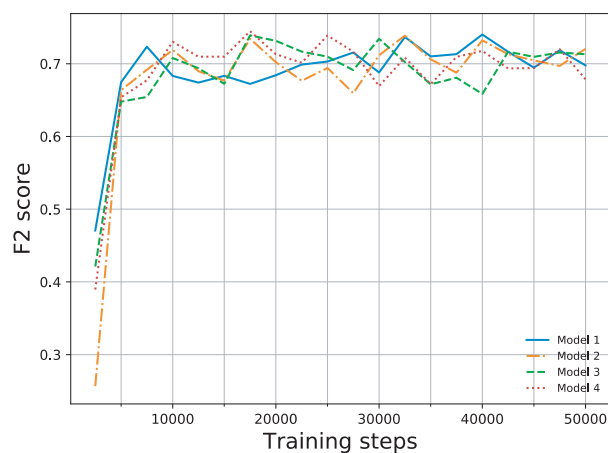


Figure 3. F2 score obtained by each of the four models considered over training steps on the validation dataset. All models converged during the first 50 000 training steps.

threshold determines which of the predicted bounding boxes are considered as final detections. Average precision allows comparison between different algorithms without the need to specify this threshold. Since our goal was to compare the algorithm performance with human performance and humans did not provide a confidence score for their detections, we could not use this metric.

The training process is stochastic due to the stochastic gradient descent algorithm used for optimization of neural network weights. We ran the CNN four times to explore how stable the algorithm output is with respect to the stochastic training process. Neural networks models are commonly run as many times as time and availability of computational resources allow. Each of the models ran for 50,000 training steps (i.e. the number of times the weights were updated by the gradient descent algorithm) on the training dataset and the performance was evaluated on the validation dataset every 2500 training steps (Figure 3). All the models reached a plateau in F2 score after around 10,000 training steps on the validation dataset. For each of the models we chose the weights obtained at the number of training steps that gave the best performance on the validation dataset.

Results

Human detection accuracy compared with CNN performance

The results show that overall for the CNN in both homogeneous & heterogeneous areas as we received an F2 score of 0.75. The CNN performed better in heterogeneous areas with an F2 score of 0.778 compared to 0.73 in homogeneous areas. The human annotator median F2 score was 0.78 and performance was better in homogeneous areas – 0.80 compared to 0.766 in heterogeneous areas. These results show that the CNN performed with high comparable accuracy compared to human detection capabilities. Visualization of one of the model detections is shown in Figure 4.

Testing detection under different image conditions

To test the applicability of the trained CNN model on an elephant population outside of our study area we test, without further training, on a known elephant population in the Maasai Mara in Kenya (Figure 5). The image covers 0.3 km² in which 32 elephants were identified. The CNN managed to detect more than half the elephants in this image (18 true positives) and the resulting F2 score was 0.57. Figure 5 provides visualization of some example CNN detections.

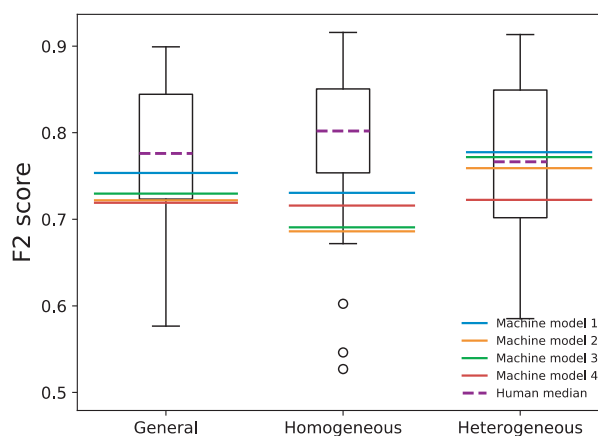


Figure 4. Results of human annotation compared with CNN detection for all images (general) and in homogenous and heterogeneous areas. The boxplots show the human results (circles represent outliers); lines are results from the four different CNN models. [Correction added on 9 July 2021, after first online publication: Figure caption has been modified.]

Discussion

Our results demonstrate, for the first time, that it is possible to automate detection of African elephants in very-high-resolution satellite imagery in both heterogeneous and homogeneous backgrounds using deep learning. We have automated the detection of elephants with as high accuracy as human detection capabilities. For a number of species remote sensing via satellite imagery is already a viable monitoring technique. However, the resolution required to identify individuals is not yet available for the vast majority of species but can be used to identify aggregations or environmental signs of presence, for example, guano stain, shadow, mounds from burrows and nests (Barber-Meyer et al., 2007; Fretwell et al., 2012; Fretwell & Trathan, 2009; Hughes et al., 2011; LaRue et al., 2014; Löffler & Margules, 1980; Lynch & LaRue, 2014; Velasco, 2009). Fortunately, a new constellation of six satellites from Maxar, Worldview Legion, will launch in 2021 that will provide imagery for the same location more than 15 times per day at 31 cm resolution. This constellation will have a tropical circle mid-inclined orbit, rather than polar orbit and will broaden the range of species that can be detected and increase detection area.

Previous studies have largely focused on marine species due both to their inaccessibility via other monitoring techniques and the high contrast of their bodies against mainly homogenous backgrounds (Barber-Meyer et al., 2007; Bowler et al., 2019; Cubaynes et al., 2018; Fretwell et al., 2017; Fretwell et al., 2014; Guirado et al., 2019; LaRue et al., 2011; LaRue & Stapleton, 2018; LaRue et al., 2015; McMahon et al., 2014; Stapleton et al., 2014). The advantages of using satellite imagery are numerous. Large

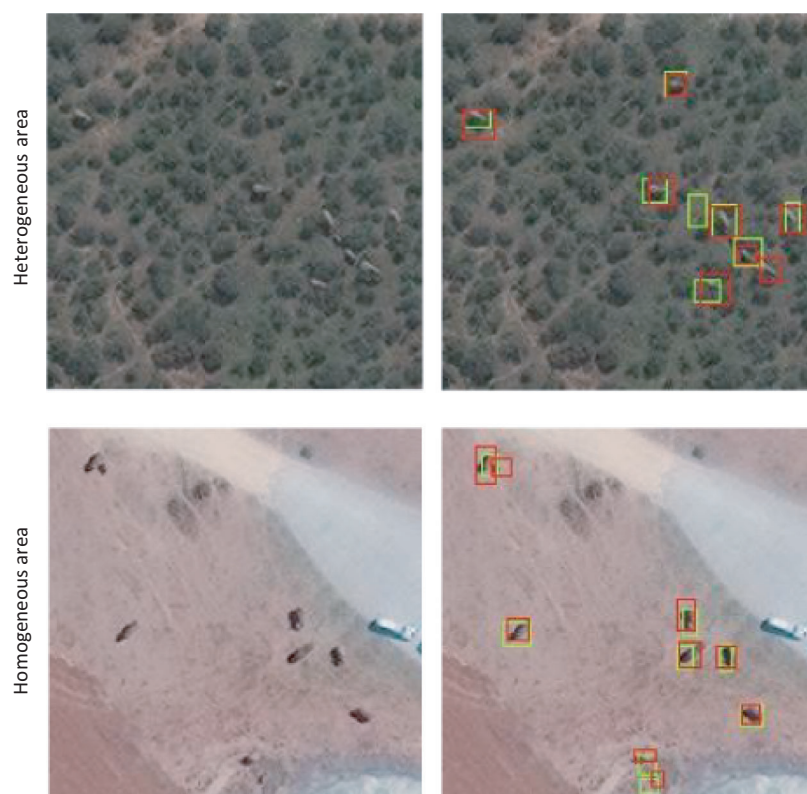


Figure 5. CNN detections: The images on the left are the raw images and images on the right are CNN detections (green boxes) and ground truth labels (red boxes). Satellite image (c) 2020 Maxar Technologies.

areas can be covered in one pass, reducing the risk of double counting and eliminating the need for repeat surveys at short intervals. Satellite monitoring is an unobtrusive technique requiring no ground presence, and thus eliminating the risk of disturbing species, or of concern for human safety during data collection. Previously inaccessible areas are rendered accessible, and cross-border areas – often crucial to conservation planning – can be surveyed without the often time-consuming and bureaucratically problematic requirements of terrestrial permits.

One challenge with satellite monitoring is the high cost of commercial satellite imagery. Worldview-3 costs \$17.50 per km² for archive imagery and tasking new imagery costs \$27.50 per km², with a minimum order of 100 km² (2020 pricing). Another key challenge is processing the large quantity of imagery generated. However, expediting identification of species by automating detection can allow for large-scale application of satellite-based wildlife surveying (LaRue et al., 2015; Torney et al., 2019). A detection process that would formally have taken weeks can thus be completed in a matter of hours. Furthermore, observer variability means errors in human-labelled datasets are inconsistently biased while in contrast, false negatives and false positives in deep learning

algorithms are consistent and can be rectified by systematically improving models. The use of Convolution Neural Networks (CNN) to automate detection of wildlife has been successfully applied on imagery from a variety of sensors including, UAVs (Bowley et al., 2018; Christie et al., 2016; Gray et al., 2018; Kellenberger et al., 2019; Kellenberger et al., 2018; Mairea et al., 2013), manned aircraft (Borowicz et al., 2019; Eikelboom et al., 2019; Maire et al., 2015; Sharma et al., 2019; Torney et al., 2019), multi-beam imaging sonar (Toshihiro et al., 2019) and camera trap imagery (Miao et al., 2019; Schneider et al., 2020; Schneider et al., 2018; Willi et al., 2018). To the best of our knowledge, only three studies have applied a CNN to satellite imagery in the case of albatross (Bowler et al., 2019), whales (Borowicz et al., 2019; Guirado et al., 2019) and pack-ice seals (Gonçalves et al., 2020).

Automating detection is becoming easier as off-the-shelf object detection tools are increasingly accessible to non-experts; however, the biggest obstacle is obtaining sufficiently large training datasets. Crowdsourced labelling platforms, e.g. Zooniverse [<https://www.zooniverse.org/>], Amazon Mechanical Turk [<https://www.mturk.com/>] can help in the creation of these bespoke training datasets using the ‘Wisdom of the crowd’ (Kao et al., 2018; Mierswa, 2016).

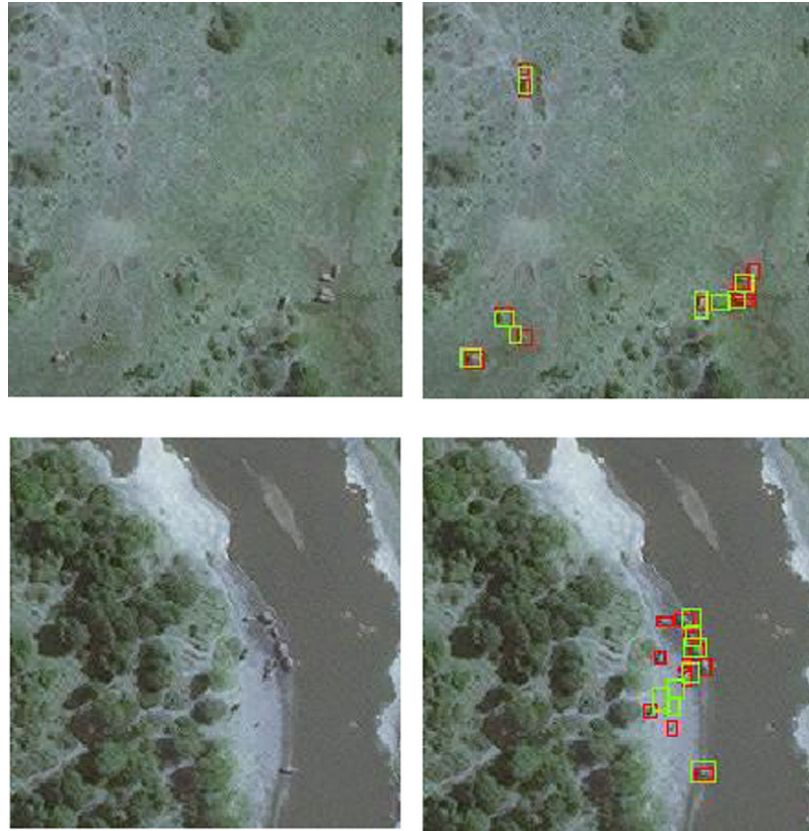


Figure 6. Example of CNN detections in Maasai Mara, Kenya from Geoeye-1 Satellite. Raw images on left and images with CNN detections (green boxes) and ground truth labels (red boxes) on right. Satellite image (c)2020 Maxar Technologies.

Our study shows the applicability of this monitoring technique to the case of the African Elephant and demonstrate, for the first time, that it is possible to automate detection of African elephants in very-high-resolution satellite imagery in both heterogeneous and homogeneous backgrounds using deep learning. Not only is the detection accuracy we achieve for elephants as high as that of humans, but there is less variation in the consistency of detection for the CNN compared to human detection performance (as shown in Figure 6).

In addition, we show that it is possible to generalize detection to elephant populations outside of the site of training data. The generalizability of the CNN is promising, as a small amount of training data from this locality or satellite would further increase accuracy. Elephant calves were accurately detected, despite their absence in the training dataset.

Areas of future research to expand this technique include testing whether performance improvements for detecting elephants can be achieved by including the near infrared band and testing to discover for which other species this is already a viable monitoring technique. In our study site vegetation cover was not heavy enough to

hinder identification but this is a common challenge when using aerial surveying techniques and would be an obstacle for forest elephants or elephants in other sites.

More broadly, deep learning methods for detecting small objects can be further improved (Cao et al., 2019; Pang et al., 2019) and large training datasets containing images of wildlife from an aerial perspective should be developed. If satellite monitoring is applied at scale then developing methods to ensure standardized and occasional ground-truthing will be required to ensure image interpretation is accurate (LaRue et al., 2017). Using high resolution satellite imagery as a wildlife surveying tool will inevitably increase in the future as image resolution improves and costs fall. Developing automated detection tools to enable larger scale application of this wildlife monitoring technique is highly valuable as satellite image surveying capabilities expand.

Acknowledgements

We greatly appreciate the generous support by the DigitalGlobe Foundation (Maxar Technologies) and European Space Agency for awarding the image grants to support

this work without which this research would not have been possible. We are also grateful to Hexagon Geospatial for providing a complementary license to ERDAS Imagine which enabled us to process the satellite imagery. We are grateful to all the human volunteer labellers for taking the time to label the images and giving us a point of comparison to the CNN. ID is grateful for a bequest to the Wildlife Conservation Research Unit, University of Oxford from the Ralph Mistler Trust which supported her to carry out this research. SR is grateful for funding from the LLOYD's Register Foundation through the Alan Turing Institute's Data Centric Engineering programme. We are grateful to Sofia Minano Gonzalez and Hannah Cubaynes for their valuable comments on the manuscript.

Authors' contributions

I.D. conceived the idea of the paper and designed the methodology with oversight from T.W.; I.D. acquired the image data, processed, analysed and labelled the satellite images; O.I. ran the code in Tensorflow with input from S.R.; I.D. and O.I. organized and analysed the human labels. I.D. led the writing of the manuscript with input from O.I., S.R., D.M. & T.W. All authors contributed critically to the drafts and gave final approval for publication.

Data availability statement

We are not at liberty to share the raw satellite imagery due to the restrictions defined in the contract with Maxar Technologies and the European Space Agency under which the image grants were awarded.

References

- Abileah, R. (2002) Marine mammal census using space satellite imagery. *U.S. Navy Journal of Underwater Acoustics*, **52**.
- Barber-Meyer, S.M., Kooyman, G.L. & Ponganis, P.J. (2007) Estimating the relative abundance of emperor penguins at inaccessible colonies using satellite imagery. *Polar Biology*, **30** (12), 1565–1570.
- Barnes, R., Beardsley, K., Micheltore, F., Barnes, K.L., Alers, M.P. & Blom, A. (1997) Estimating forest elephant numbers with dung counts and a geographic information system. *The Journal of Wildlife Management*, **61**(4), 1384–1393.
- Barnosky, A.D., Brown, J.H., Daily, G.C., Dirzo, R., Ehrlich, A.H. & Ehrlich, P.R. et al. (2014) Introducing the Scientific Consensus on Maintaining Humanity's Life Support Systems in the 21st Century: Information for Policy Makers. *The Anthropocene Review*, **1**(1), 78–109.
- 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.
- Bowler, E. et al. (2019) Using deep learning to count albatrosses from space: Assessing results in light of ground truth uncertainty. *Remote Sensing*, **12**.
- Bowley, C. et al. Detecting Wildlife in Unmanned Aerial Systems Imagery Using Convolutional Neural Networks Trained with an Automated Feedback Loop. *Lecture Notes in Computer Science*. Vol. 10860. 2018: Springer, Cham.
- Bruijning, M. et al. (2018) trackdem: Automated particle tracking to obtain population counts and size distributions from videos in r. *Methods in Ecology and Evolution*, **9**(4), 965–973.
- Buma, W.G. & Lee, S.-I. (2019) Multispectral image-based estimation of drought patterns and intensity around Lake Chad, Africa. *Remote Sensing*, **11**(21).
- Cao, C., Wang, B., Zhang, W., Zeng, X., Yan, X. & Feng, Z. et al. (2019) An Improved Faster R-CNN for Small Object Detection. *IEEE Access*, **7**, 106838–106846.
- Cardinale, B. (2012) Impacts of Biodiversity Loss. *Science*, **336**, 552–553.
- Caughley, G., Sinclair, R. & Scott-Kemmis, D. (1976) Experiments in aerial survey. *The Journal of Wildlife Management*, **40**, 290–300.
- Chabot, D. & Bird, D.M. (2015) Wildlife research and management methods in the 21st century: Where do unmanned aircraft fit in? *Journal of Unmanned Vehicle Systems*, **3**(4), 137–155.
- Christie, K.S., Gilbert, S.L., Brown, C.L., Hatfield, M. & Hanson, L. (2016) Unmanned aircraft systems in wildlife research: current and future applications of transformative technology. *Frontiers in Ecology and the Environment*, **14**(5), 241–251.
- Cubaynes, H.C., Fretwell, P.T., Bamford, C., Gerrish, L. & Jackson, J.A. et al. (2018) Whales from space: Four mysticete species described using new VHR satellite imagery. *Marine Mammal Science*, **35**(2), 466–491.
- Dascalu, A. & David, E.O. (2019) Skin cancer detection by deep learning and sound analysis algorithms: A prospective clinical study of an elementary dermoscope. *EBioMedicine*, **43**, 107–113.
- de Carvalho Júnior, O., Guimarães, R.F., Silva, C.R. & Gomes, R.A. (2015) Standardized Time-Series and Interannual Phenological Deviation: New Techniques for Burned-Area Detection Using Long-Term MODIS-NBR Dataset. *Remote Sensing*, **7**(6), 6950–6985.
- Du Toit, J.C.O. & O'Connor, T.G. (2014) Changes in rainfall pattern in the eastern Karoo, South Africa, over the past 123 years. *Water SA*, **40**(3).
- Dutta, A. & Zisserman, A. (2019). The VIA Annotation Software for Images, Audio and Video. in *ACM International Conference on Multimedia* (2019) Nice, France: ACM.

- Eikelboom, J.A.J., Wind, J., van de Ven, E., Kenana, L.M., Schroder, B. & de Knecht, H.J. et al. (2019) Improving the precision and accuracy of animal population estimates with aerial image object detection. *Methods in Ecology and Evolution*, **10**(11), 1875–1887.
- Emmanuel, H., Di Vittorio, M., Barnes, R.F., Guenda, W. & Luiselli, L. (2017) Detection of interannual population trends in seven herbivores from a West African savannah: a comparison between dung counts and direct counts of individuals. *African Journal of Ecology*, **55**, 609–617.
- Ferreira, A.C., Silva, L.R., Renna, F., Brandl, H.B., Renoult, J.P. & Farine, D.R. et al. (2020) Deep learning-based methods for individual recognition in small birds. *Methods in Ecology and Evolution*.
- Fretwell, P.T., LaRue, M.A., Morin, P., Kooyman, G.L., Wienecke, B. & Ratcliffe, N. et al. (2012) An emperor penguin population estimate: the first global, synoptic survey of a species from space. *PLoS One*, **7**(4), e33751.
- Fretwell, P.T., Scofield, P. & Phillips, R.A. (2017) Using super-high resolution satellite imagery to census threatened albatrosses. *Ibis*, **159**(3), 481–490.
- Fretwell, P.T., Staniland, I.J. & Forcada, J. (2014) Whales from space: counting southern right whales by satellite. *PLoS One*, **9**(2), e88655.
- Fretwell, P.T. & Trathan, P.N. (2009) Penguins from space: faecal stains reveal the location of emperor penguin colonies. *Global Ecology and Biogeography*, **18**, 543–552.
- Fullman, T., Kiker, G.A., Gaylard, A., Southworth, J., Waylen, P. & Kerley, G.I. et al. (2017) Elephants respond to resource trade-offs in an aseasonal system through daily and annual variability in resource selection. *KOEDOE - African Protected Area Conservation and Science*, 1–21.
- Gara, T.E.A. (2016) Understanding the effect of landscape fragmentation and vegetation productivity on elephant habitat utilization in Amboseli ecosystem, Kenya. *African Journal of Ecology*, **55**, 860–871.
- Ginosar, S., Haas, D., Brown, T. & Malik, J. (2014). Detecting People in Cubist Art. in Computer Vision ECCV 2014 Workshops. Springer, Cham.
- Gonçalves, B.C., Spitzbart, B. & Lynch, H.J. (2020) SealNet: A fully-automated pack-ice seal detection pipeline for sub-meter satellite imagery. *Remote Sensing of Environment*, **239**.
- Gray, P.C., Fleishman, A.B., Klein, D.J., McKown, M.W., Bézy, V.S. & Lohmann, K.J. et al. (2018) A Convolutional neural network for detecting sea turtles in drone imagery. *Methods in Ecology and Evolution*.
- Guirado, E., Tabik, S., Rivas, M.L., Alcaraz-Segura, D. & Herrera, F. (2019) Whale counting in satellite and aerial images with deep learning. *Scientific Reports*, **9**(1), 14259.
- Gulshan, V., Peng, L., Coram, M., Stumpe, M.C., Wu, D. & Narayanaswamy, A. et al. (2016) Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*, **316**(22), 2402–2410.
- Hollings, T., Burgman, M., van Andel, M., Gilbert, M., Robinson, T. & Robinson, A. (2018) How do you find the green sheep? A critical review of the use of remotely sensed imagery to detect and count animals. *Methods in Ecology and Evolution*, **9**(4), 881–892.
- Hordiiuk, D., Oliinyk, I., Hnatushenko, V. & Maksymov, K. (2019) Semantic Segmentation for Ships Detection from Satellite Imagery, in IEEE 39th International Conference on Electronics and Nanotechnology (ELNANO). 2019. pp. 454–457.
- Huang, J., Rathod, V., Sun, C., Zhu, M., Korattikara, A. & Fathi, A. et al. (2017) Speed/Accuracy Trade-Offs for Modern Convolutional Object Detectors, in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017. pp. 3296–3297.
- Hughes, B.J., Martin, G.R. & Reynolds, S.J. (2011) The use of Google EarthTMsatellite imagery to detect the nests of masked boobies *Sula dactylatra*. *Wildlife Biology*, **17**(2), 210–216.
- Jachmann, H. (2002) Comparison of Aerial Counts with Ground Counts for Large African Herbivores. *British Ecological Society*, **39**(5), 841–852.
- Kakembo, V., Smith, J. & Kerley, G. (2015) A Temporal Analysis of Elephant-Induced Thicket Degradation in Addo Elephant National Park, Eastern Cape, South Africa. *Rangeland Ecology & Management*, **68**(6), 461–469.
- Kao, A.B., Berdahl, A.M., Hartnett, A.T., Lutz, M.J., Bak-Coleman, J.B. & Ioannou, C.C. et al. (2018) Counteracting estimation bias and social influence to improve the wisdom of crowds. *Journal of the Royal Society, Interface*, **15**(141).
- Kellenberger, B., Marcos, D., Lobry, S. & Tuia, D. (2019) Half a percent of labels is enough: efficient animal detection in UAV imagery using deep CNNs and active learning. *IEEE Transactions on Geoscience and Remote Sensing*, **57**(12), 9524–9533.
- Kellenberger, B., Marcos, D. & Tuia, D. (2018) Detecting mammals in UAV images: Best practices to address a substantially imbalanced dataset with deep learning. *Remote Sensing of Environment*, **216**, 139–153.
- Kiage, L.M., Liu, K.B., Walker, N.D., Lam, N. & Huh, O.K. (2007) Recent land-cover/use change associated with land degradation in the Lake Baringo catchment, Kenya, East Africa: evidence from Landsat TM and ETM+. *International Journal of Remote Sensing*, **28**(19), 4285–4309.
- Koneff, M.D., Royle, J.A., Otto, M.C., Wortham, J.S. & Bidwell, J.K. (2008) A double-observer method to estimate detection rate during aerial waterfowl surveys. *The Journal of Wildlife Management*, **72**, 1641–1649.
- Krizhevsky, A., Sutskever, I. & Hinton, G. ImageNet Classification with Deep Convolutional Neural Networks. in Advances in Neural Information Processing Systems 2012.
- Kusimi, J.M. (2008) Assessing land use and land cover change in the Wassa West District of Ghana using remote sensing. *GeoJournal*, **71**(4), 249–259.

- LaRue, M.A., Lynch, H.J., Lyver, P.O., Barton, K., Ainley, D.G. & Pollard, A. et al. (2014) A method for estimating colony sizes of Adélie penguins using remote sensing imagery. *Polar Biology*, **37**(4), 507–517.
- LaRue, M.A., Rotella, J.J., Garrott, R.A., Siniff, D.B., Ainley, D.G. & Stauffer, G.E. (2011) Satellite imagery can be used to detect variation in abundance of Weddell seals (*Leptonychotes weddellii*) in Erebus Bay, Antarctica. *Polar Biology*, **34**(11), 1727–1737.
- LaRue, M.A. & Stapleton, S. (2018) Estimating the abundance of polar bears on Wrangel Island during late summer using high-resolution satellite imagery: a pilot study. *Polar Biology*, **41**(12), 2621–2626.
- LaRue, M.A., Stapleton, S. & Anderson, M. (2017) Feasibility of using high-resolution satellite imagery to assess vertebrate wildlife populations. *Conservation Biology*, **31**(1), 213–220.
- LaRue, M.A., Stapleton, S., Porter, C., Atkinson, S., Atwood, T. & Dyck, M. et al. (2015) Testing methods for using high-resolution satellite imagery to monitor polar bear abundance and distribution. *Wildlife Society Bulletin*, **39**(4), 772–779.
- LeCun, Y., Bengio, Y. & Hinton, G. (2015) Deep learning. *Nature*, **521**(7553), 436–444.
- Lin, T., Microsoft COCO: Common Objects in Context, in Computer Vision – ECCV (2014). 2014. Cham: Springer.
- Löffler, E. & Margules, C. (1980) Wombats Detected from Space. *Remote Sensing of Environment*, **9**, 47–56.
- Lu, D., Mausel, P., Brondizio, E. & Moran, E. (2010) Change detection techniques. *International Journal of Remote Sensing*, **25**(12), 2365–2401.
- Lynch, H.J. & LaRue, M.A. (2014) First global census of the Adélie Penguin. *The Auk*, **131**(4), 457–466.
- Lynch, H.J., White, R., Black, A.D. & Naveen, R. (2012) Detection, differentiation, and abundance estimation of penguin species by high-resolution satellite imagery. *Polar Biology*, **35**(6), 963–968.
- Maire, F., Alvarez, L.M. & Hodgson, A. (2015) Automating marine mammal detection in aerial images captured during wildlife surveys: A deep learning approach. 9457, 379–385.
- Maire, F., Mejias, L., Hodgson, A. & Duclos, G. (2013) Detection of Dugongs from Unmanned Aerial Vehicles. in IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
- McMahon, C.R., Howe, H., Van Den Hoff, J., Alderman, R., Broolsma, H. & Hindell, M.A. (2014) Satellites, the all-seeing eyes in the sky: counting elephant seals from space. *PLoS One*, **9**(3), e92613.
- Meng, Q. & Meentemeyer, R.K. (2011) Modeling of multi-strata forest fire severity using Landsat TM Data. *International Journal of Applied Earth Observation and Geoinformation*, **13**(1), 120–126.
- Miao, Z., Gaynor, K.M., Wang, J., Liu, Z., Muellerklein, O. & Norouzzadeh, M.S. et al. (2019) Insights and approaches using deep learning to classify wildlife. *Scientific Reports*, **9**(1), 8137.
- Mierswa, I. (2016) The Wisdom of Crowds: Best Practices for Data Prep & Machine Learning Derived from Millions of Data Science Workflows, in nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016. p. 411.
- Mulero-Pazmany, M., Jenni-Eiermann, S., Strebel, N., Sattler, T., Negro, J.J. & Tablado, Z. (2017) Unmanned aircraft systems as a new source of disturbance for wildlife: a systematic review. *PLoS One*, **12**(6), e0178448.
- Olczak, J., Fahlberg, N., Maki, A., Razavian, A.S., Jilert, A. & Stark, A. et al. (2017) Artificial intelligence for analyzing orthopedic trauma radiographs. *Acta Orthopaedica*, **88**(6), 581–586.
- Oliveira, E.R., Disperati, L., Cenci, L., Gomes Pereira, L. & Alves, F.L. (2019) Multi-index image differencing method (MINDED) for Flood Extent Estimations. *Remote Sensing*, **11**(11).
- Pang, J., Li, C., Shi, J., Xu, Z. & Feng, H. (2019) CNN: Fast tiny object detection in large-scale remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, **57**(8), 5512–5524.
- Petersen, S. et al. Using machine learning to accelerate ecological research. 2019, Deepmind.p. Blog Post Research.
- Poulsen, J.R., Koerner, S.E., Moore, S., Medjibe, V.P., Blake, S. & Clark, C.J. et al. (2017) Poaching empties critical Central African wilderness of forest elephants. *Current Biology*, **27**(4), R134–R135.
- Ren, S., He, K., Girshick, R. & Sun, J. (2017) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**(6), 1137–1149.
- Romero, J.R., Roncallo, P.F., Akkiraju, P.C., Ponzoni, I., Echenique, V.C. & Carballido, J.A. (2013) Using classification algorithms for predicting durum wheat yield in the province of Buenos Aires. *Computers and Electronics in Agriculture*, **96**, 173–179.
- Rulinda, C.M., Bijker, W. & Stein, A. (2010) Image mining for drought monitoring in eastern Africa using Meteosat SEVIRI data. *International Journal of Applied Earth Observation and Geoinformation*, **12**, S63–S68.
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S. & Ma, S. et al. (2015) ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, **115**(3), 211–252.
- Sasamal, S.K., Chaudhury, S.B., Samal, R.N. & Pattanaik, A.K. (2008) QuickBird spots flamingos off Nalabana Island, Chilika Lake, India. *International Journal of Remote Sensing*, **29**(16), 4865–4870.
- Schlossberg, S., Chase, M.J. & Griffin, C.R. (2016) Testing the Accuracy of Aerial Surveys for Large Mammals: An Experiment with African Savanna Elephants (*Loxodonta africana*). *PLoS One*, **11**(10), e0164904.
- Schmidhuber, J. (2015) Deep learning in neural networks: an overview. *Neural Netw*, **61**, 85–117.

- Schneider, S., Greenberg, S., Taylor, G.W. & Kremer, S.C. (2020) Three critical factors affecting automated image species recognition performance for camera traps. *Ecology and Evolution*, **10**(7), 3503–3517.
- Schneider, S., Taylor, G. & Kremer, S. Deep Learning Object Detection Methods for Ecological Camera Trap Data.
- Sharma, N., Blumenstein, M. & Scully-Power, P. (2018). Shark detection from aerial imagery using region-based CNN, a study.in Lecture Notes in Computer Science.
- Shiu, Y., Wrege, P.H., Keen, S. & Rowland, E.D. (2014) Large-scale automatic acoustic monitoring of African forest elephants' calls in the terrestrial acoustic recordings. *The Journal of the Acoustical Society of America*, **135**(5), 2334–2334.
- Sibanda, M. & Murwira, A. (2012) Cotton fields drive elephant habitat fragmentation in the Mid Zambezi Valley, Zimbabwe. *International Journal of Applied Earth Observation and Geoinformation*, **19**, 286–297.
- Skogen, K., Helland, H. & Kaltenborn, B. (2018) Concern about climate change, biodiversity loss, habitat degradation and landscape change: Embedded in different packages of environmental concern? *Journal for Nature Conservation*, **44**, 12–20.
- Smit, J., Pozo, R.A., Cusack, J.J., Nowak, K. & Jones, T. (2017) Using camera traps to study the age–sex structure and behaviour of crop-using elephants *Loxodonta africana* in Udzungwa Mountains National Park, Tanzania. *Oryx*, **53**(2), 368–376.
- Soltis, J., King, L., Vollrath, F. & Douglas-Hamilton, I. (2016) Accelerometers and simple algorithms identify activity budgets and body orientation in African elephants *Loxodonta africana*. *Endangered Species Research*, [cited 31, 1–12].
- Stapleton, S., LaRue, M., Lecomte, N., Atkinson, S., Garshelis, D. & Porter, C. et al. (2014) Polar bears from space: assessing satellite imagery as a tool to track Arctic wildlife. *PLoS One*, **9**(7), e101513.
- Strigl, D., Kofler, K. & Podlipnig, S. (2010) Performance and Scalability of GPU-Based Convolutional Neural Networks. Euromicro Conference on Parallel, Distributed and Network-based Processing., 317–324 in 2010 18th.
- Szegedy, C., Ioffe, S., Vanhoucke, V. & Alemi, A. (2017) Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning.in Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. Google Inc., Amphitheatre Parkway Mountain View. CA: Association for the Advancement of Artificial Intelligence, p. (AAAI-17)..
- Tambling, C., Druce, D.J., Hayward, M.W., Castley, J.G., Adendorff, J. & Kerley, G.I. (2012) Spatial and temporal changes in group dynamics and range use enable anti-predator responses in African buffalo. *Ecology*, **93**, 1297–1304.
- Thito, K., Wolski, P. & Murray-Hudson, M. (2016) Mapping inundation extent, frequency and duration in the Okavango Delta from 2001 to 2012. *African Journal of Aquatic Science*, **41**(3), 267–277.
- Torney, C.J., Lloyd-Jones, D.J., Chevallier, M., Moyer, D.C., Maliti, H.T. & Mwitia M. et al. (2019) A comparison of deep learning and citizen science techniques for counting wildlife in aerial survey images. *Methods in Ecology and Evolution*, **10**(6), 779–787.
- Toshihiro, M., Horimoto, H., Ishihara, T. & Kofuji, K. (2019) Autonomous Tracking of Sea Turtles based on Multibeam Imaging Sonar: Toward Robotic Observation of Marine Life. *Science Direct*, **52**, 86–90.
- Tzutalin. LabelImg. Github. 2015.
- van Wilgen, N.J., Goodall, V., Holness, S., Chown, S.L. & McGeoch, M.A. (2016) Rising temperatures and changing rainfall patterns in South Africa's national parks. *International Journal of Climatology*, **36**(2), 706–721.
- Velasco, M. (2009) A Quickbird's eye view on marmots, in International Institute for Geo-information science and Earth Observation.
- Weinstein, B.G. (2018) A computer vision for animal ecology. *Journal of Animal Ecology*, **87**(3), 533–545.
- Wickler, W. & Seibt, U. (1997) Aimed Object-throwing by a Wild African Elephant in an interspecific encounter. *Ethology*, **103**(5), 365–368.
- Willi, M., Pitman, R.T., Cardoso, A.W., Locke, C., Swanson, A. & Boyer, A. et al. (2018) Identifying animal species in camera trap images using deep learning and citizen science. *Methods in Ecology and Evolution*, **10**(1), 80–91.
- Witharana, C., LaRue, M.A. & Lynch, H.J. (2016) Benchmarking of data fusion algorithms in support of earth observation based Antarctic wildlife monitoring. *ISPRS Journal of Photogrammetry and Remote Sensing*, **113**, 124–143.
- Xue, Y., Wang, T. & Skidmore, A.K. (2017) Automatic Counting of Large Mammals from Very-high Resolution Panchromatic Satellite Imagery. *Remote Sensing*, **9**(9).
- Yang, Z., Wang, T., Skidmore, A.K., de Leeuw, J., Said, M.Y. & Freer, J. (2014) Spotting East African mammals in open savannah from space. *PLoS One*, **9**(12), e115989.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Appendix S1. Image list.

Appendix S2. Comparison of human and machine results on individual images.