Received: 15 October 2018 | Accepted: 22 January 2019

DOI: 10.1111/2041-210X.13165

RESEARCH ARTICLE

Methods in Ecology and Evolution BRITISH ECOLOGICAL SOCIETY

A comparison of deep learning and citizen science techniques for counting wildlife in aerial survey images

Colin J. Torney[1*] | David J. Lloyd-Jones[2*] | Mark Chevallier[1] | David C. Moyer[3] | Honori T. Maliti[4] | Machoke Mwita[4] | Edward M. Kohi[4] | Grant C. Hopcraft[5]

[1]School of Mathematics and Statistics, University of Glasgow, Glasgow, UK

[2]FitzPatrick Institute of African Ornithology, DST-NRF Centre of Excellence, University of Cape Town, Rondebosch, South Africa

[3]Integrated Research Center, The Field Museum of Natural History, Chicago, Illinois

[4]Tanzania Wildlife Research Institute, Arusha, Tanzania

[5]Institute of Biodiversity, Animal Health and Comparative Medicine, University of Glasgow, Glasgow, UK

Correspondence
Colin J. Torney
Email: colin.j.torney@gmail.com

Funding information
European Union Horizon 2020, Grant/Award Number: 641918; James S. McDonnell Foundation

Handling Editor: Rachel McCrea

Abstract

1. Fast and accurate estimates of wildlife abundance are an essential component of efforts to conserve ecosystems in the face of rapid environmental change. A widely used method for estimating species abundance involves flying aerial transects, taking photographs, counting animals within the images and then inferring total population size based on a statistical estimate of species density in the region. The intermediate task of manually counting the aerial images is highly labour intensive and is often the limiting step in making a population estimate.
2. Here, we assess the use of two novel approaches to perform this task by deploying both citizen scientists and deep learning to count aerial images of the 2015 survey of wildebeest (*Connochaetes taurinus*) in Serengeti National Park, Tanzania.
3. Through the use of the online platform Zooniverse, we collected multiple non-expert counts by citizen scientists and used three different aggregation methods to obtain a single count for the survey images. We also counted the images by developing a bespoke deep learning method via the use of a convolutional neural network. The results of both approaches were then compared.
4. After filtering of the citizen science counts, both approaches provided highly accurate total estimates. The deep learning method was far faster and appears to be a more reliable and predictable approach; however, we note that citizen science volunteers played an important role when creating training data for the algorithm. Notably, our results show that accurate, species-specific, automated counting of aerial wildlife images is now possible.

KEYWORDS

citizen science, conservation, deep learning, monitoring, population ecology, surveys

1 | INTRODUCTION

Estimating the abundance of animal species is essential for ecologists, conservationists and wildlife managers world-wide. Measuring population abundance enables the early detection of population declines (caused by disease, over-harvesting or changing patterns of land-use), or population increases and expansions; therefore, it is a precursor for adaptive management and conservation strategies (Walters, 1986). Repeated measurements of population abundance also provide insight into the key factors that regulate natural

*These authors contributed equally to this work.

populations (Turchin, 1999), a means to determine their vital rates (Mduma, Sinclair, & Hilborn, 1999), and are an essential requirement for validating theoretical models of species interactions.

The challenge of detecting and responding to changes in animal abundance is especially acute in the case of migratory species (Harris, Thirgood, Hopcraft, Cromsigt, & Berger, 2009; Singh & Milner-Gulland, 2011). Estimating population sizes of migratory species is a vital, but logistically challenging task. Localised environmental disturbances in large geographical areas are often hard to detect, while the fact that migrations span national and regional borders, means implementing protection strategies typically involves substantial time to coordinate (Lovejoy, Sallaberry, Senner, & Tarak, 1987; Wilcove & Wikelski, 2008).

The Serengeti National Park in Tanzania is known for the iconic migration of *c*. 1.3 million blue wildebeest *Connochaetes taurinus* and 250,000 common zebra *Equus quagga*. This is the largest terrestrial migration of animals on Earth (Thirgood et al., 2004) and their annual movement alters every biological process in the ecosystem, from soil nutrient cycles to the diversity of insects, birds and carnivores, to the balance of trees and grass (Estes, 2014; Holdo, Holt, Sinclair, Godley, & Thirgood, 2011; McNaughton, 1985; Subalusky, Dutton, Rosi, & Post, 2017), as well as providing vital ecosystem services to human communities around the park (Sinclair, Metzger, Mduma, & Fryxell, 2015). Without the annual migratory cycle, there would be fundamental changes in the ecology of the region and much of its biodiversity would decline (Dobson et al., 2010; Holdo, Fryxell, Sinclair, Dobson, & Holt, 2011). The long-term population trend of the wildebeest (see Figure 1) is closely tied to levels of poaching, disease, climate change and human perturbations. Therefore, estimating wildebeest abundance is perhaps the most important metric of the ecosystem's health (Estes, 2014).

The standard approach to counting the wildebeest population is to fly transects over the herds in March, April or May (Campbell & Borner, 1995; Norton-Griffiths, 1973) while the bulk of the wildebeest are on the short grass plains in the south-east of Serengeti and the Ngorongoro conservation area, before the migration moves into the woodland areas of the western Serengeti. As with many aggregated species, instead of performing in air counts, nadir geo-referenced aerial photographs are taken of the survey area at fixed intervals from an aircraft flown as close as possible to a constant speed. The next stage of the process is then to identify and count all wildebeest within each image. This process of manually counting each image is a labour-intensive process that typically takes three or four skilled counters *c*. 3–6 weeks (Torney et al., 2016). Automating this aspect of the survey would have two major advantages. First, it would remove a bottleneck in running the survey. Beyond the actual

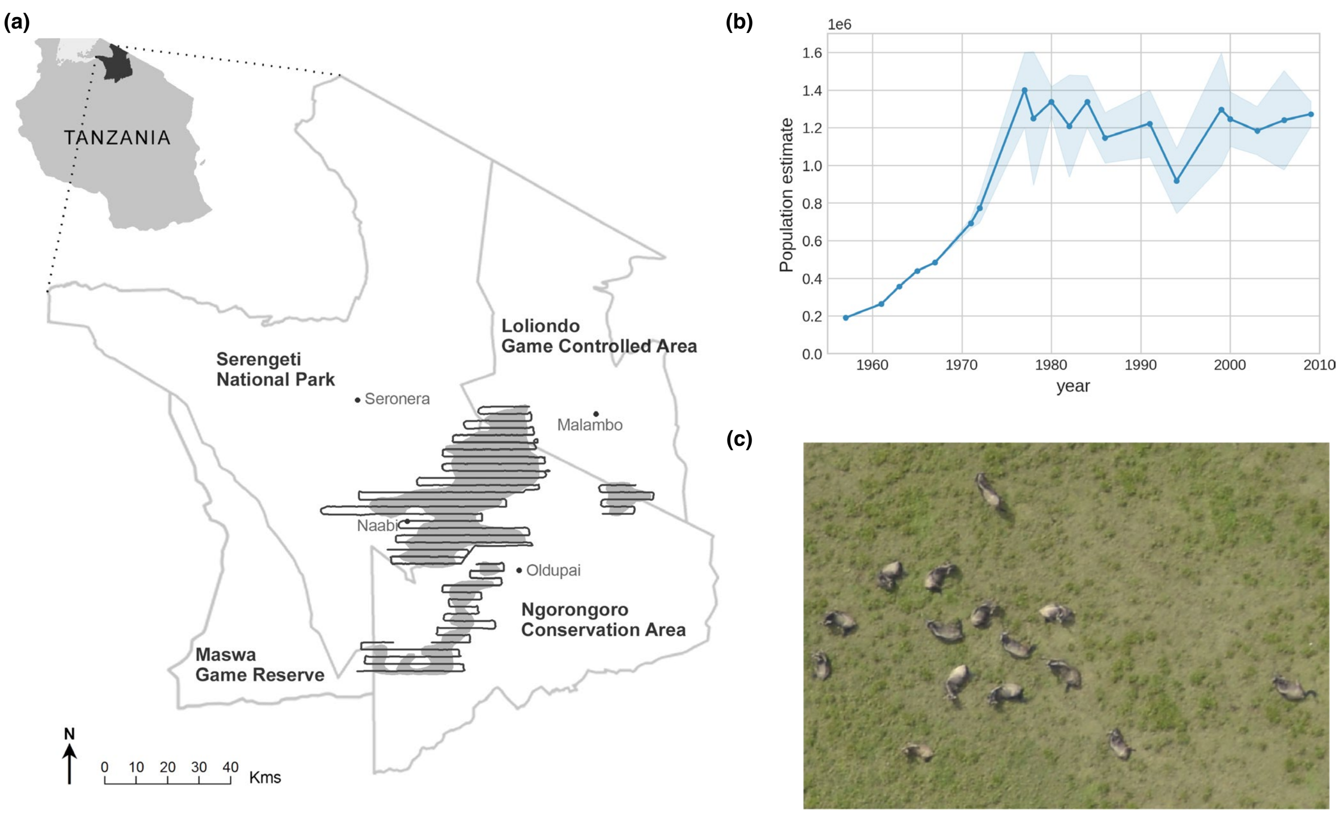

FIGURE 1 The Serengeti wildebeest census. (a) Map of the region showing the approximate distribution of the herds (shaded) and the transects flown during the 2015 census. (b) Wildebeest population estimates over time. Shaded region indicates the standard error of the estimate. Since the 1970s, estimates have been made using the Jolly II method (Jolly, 1969) with data collected using vertical aerial photography and manual image counts. Images from a census conducted in 2018 are currently being manually processed. Infrequent counts with large errors highlight the need for new approaches. (c) An example section of a 2015 survey image showing a group of wildebeest

counting time, there is often a considerable delay in scheduling the counting process as it involves multiple wildlife professionals to undertake. Second, removing this time-consuming job would relieve a significant burden on the organisations involved, freeing conservation professionals to focus on other tasks. Two novel methods can potentially replace the use of manual counts by experts, the deployment of citizen scientists and the use of automated object detection algorithms. In this work, we deploy both approaches and evaluate the performance and merits of each.

1.1 | Citizen science and the wisdom of crowds

It has long been noted that multiple non-expert individuals can be as accurate as a single expert for certain tasks if their estimates are appropriately aggregated (Condorcet, 1976; Galton, 1907). This phenomenon represents collective intelligence in its purest form, or as it's commonly known 'the wisdom of crowds' (Surowiecki, 2005) and in effect means that as more individuals estimate some quantity of interest, then an appropriate aggregate quantity (Kao et al., 2018) derived from these estimates will converge on the true value.

The wisdom of crowds is the basis for many attempts to harness the collective power of citizen scientists. The key idea is that through online platforms, such as Zooniverse (Simpson, Page, & De Roure, 2014), scientists can outsource tasks to non-experts and by aggregating multiple responses obtain usable, reliable data. Citizen science has been used in multiple domains from protein folding (Dill & MacCallum, 2012) to astronomy (Lintott et al., 2008), and appears to be growing as a tool for ecologists and conservationists (Ellwood, Crimmins, & Miller-Rushing, 2017; Swanson, Kosmala, Lintott, & Packer, 2016), where it has the major advantage of not only performing scientific analysis of data but also engaging the public with wildlife conservation (Forrester et al., 2017). However, despite the growth in the use and awareness of citizen science approaches, there is still some scepticism about the reliability of unpaid and often anonymous volunteers (see (Kosmala, Wiggins, Swanson, & Simmons, 2016; Sauermann & Franzoni, 2015) for a review and discussion of these issues and potential mitigation strategies).

1.2 | Automated computer vision

Another potential approach to replacing dedicated professional counters is to use machine learning algorithms. Computer vision and machine learning are increasingly becoming essential components of the ecologist's toolbox (Bruijning, Visser, Hallmann, & Jongejans, 2018; Christin, Hervet, & Lecomte, 2018; Dell et al., 2014; Mac Aodha et al., 2018; Valletta, Torney, Kings, Thornton, & Madden, 2017; Weinstein, 2018) and have been applied previously to the task of counting aerial images of animals (Bajzak & Piatt, 1990; Chabot, Dillon, & Francis, 2018; Laliberte & Ripple, 2003; McNeill, Barton, Lyver, & Pairman, 2011; Rey, Volpi, Joost, & Tuia, 2017; Xue, Wang, & Skidmore, 2017; Yang et al., 2014) including the Serengeti wildebeest population (Torney et al., 2016).

While attempts to automate the classification and/or localisation of objects within images have been ongoing for decades, recently a combination of advances in machine learning, increased parallel computing power provided by graphical processing units (GPUs) and accessibility of image training datasets, such as the COCO dataset (Lin et al., 2014), have led to rapid improvements in the performance of multilayer deep convolutional neural networks (DCNNs). These multilayer neural networks are a form of deep learning and are distinct to traditional machine learning approaches to computer vision in that no hand-crafted features are required; instead, the convolutional layers extract relevant features directly from the training data. For image classification tasks, DCNNs achieved accuracy levels that match the ability of humans a number of years ago (Szegedy et al., 2015). Computationally efficient object detection is a more difficult task as it effectively involves multiple classifications of different regions within an image. Recently, a number of specialised object detection networks have been developed that either use a two-stage process of proposing regions then classifying them (Girshick, 2015; Ren, He, Girshick, & Sun, 2015) or a single pass through the network to predict object classes and their coordinates (Liu et al., 2016; Redmon, Divvala, Girshick, & Farhadi, 2016; Redmon & Farhadi, 2017). In this work, we evaluate the performance of the single-pass DCNN architecture proposed in (Redmon et al., 2016) and iteratively refined in (Redmon & Farhadi, 2017, 2018), named YOLO which stands for *you only look once*, referring to the fact only a single pass through the network is required.

2 | MATERIALS AND METHODS

2.1 | Aerial surveys

The 2015 Serengeti wildebeest count was conducted between 23 April and 2 May over the eastern and southern plains of Serengeti National Park, Ngorongoro Conservation Area, Loliondo Game Controlled Area and Maswa Game Reserve. A Cessna C182 aircraft was used to conduct the survey, with photographs taken using a NIKON D800 through a 35 mm Nikor Lens. The camera was mounted in a port in the floor of the aircraft and images were manually triggered at the start of each transect to be collected automatically every 10 s.

Reconnaissance flights over several days prior to the count identified the distribution of the migratory herd and from these flights the herd distribution was mapped, and a survey frame identified. When the distribution was optimal, 10.3 hours of photographic sampling flights were flown along east-west transects on 30 April and 2 May covering a straight-line distance of 2,040 km.

During the count, flight target altitude along transects was 700 ft (213 m) above the ground. This was an optimal height to both maximise image resolution but not startle the wildebeest into running from the sound of the aircraft engine. Ground speed was maintained as closely as possible to 100 knots (185 km/h). A total of 1584 georeferenced images were taken with a resolution of 7,360 × 4,912 pixels.

2.2 | Zooniverse image counts

The citizen science approach was facilitated by development of a wildebeest counting website using the Zooniverse platform (Simpson et al., 2014). Images were first filtered to remove those that were known to definitely not contain any wildebeest. This step was taken to reduce the number of images that needed to be uploaded and, more importantly, reduce the number of empty images that citizen counters needed to process.

Following initial trials on the website in August 2015, it was determined that volunteers would struggle to count entire aerial images due to their large size and high resolution. Our solution was to split each aerial image into 12 equal-sized tiles. The images were uploaded to the Serengeti Wildebeest Count project on the Zooniverse platform, which included an information page, a *Field Guide* to help with the identification of wildebeest and other animals, and the actual display of images where users could click on the images to indicate where they thought a wildebeest was present. The pixel locations of each click were then recorded.

A total of 9,870 images were counted by 2,212 volunteers between 10 May and 31 May 2017. Anyone could visit the website and count wildebeest, and each image was counted by 15 different volunteers. Once an image was counted 15 times, it was retired and the overall project progress was displayed on a statistics bar on the home page. Once all images were retired the classification data was downloaded. The data included the number of wildebeest counted by each user, their username (unregistered users were given a random username) and the pixel location of each of their identifications. Prior to analysis, any count data made using early versions of the counting interface or collected by either developers or citizen scientists during testing of the interface functionality was removed.

2.3 | Implementation of object detection algorithm

To automate the image counts, we implemented the YOLOv3 (Redmon & Farhadi, 2018) object detector using the open source deep learning packages Keras (Chollet et al., 2015) and TensorFlow (Abadi et al., 2016). The implementation of the algorithm followed three main steps.

Firstly, we generated a training dataset by selecting 500 of the survey images at random to be used exclusively for training. Images were tiled into 864 × 864 subimages and then passed though a version of the YOLO DCNN using pretrained weights from the COCO dataset provided by (Redmon & Farhadi, 2018). This process created a list of the locations of potential objects in each image. As a first pass, these results were filtered by discarding any object detections that did not correspond to an identification from the Zooniverse data. After this initial filter, the bounding boxes were manually checked and corrected for each of the 20,000 training images (500 full size images were each divided into 40, 864 × 864 training images).

With this training set, we next made several minor modifications to the YOLOv3 architecture. YOLOv3 employs nine predefined object shapes, termed anchor boxes, as initial estimates for object bounding box heights and widths. As there was less variation in our target objects, we reduced the number of anchor boxes used from nine to three, and replaced the dimensions of the three anchor boxes to match the training dataset. We next removed all but the final scale boxes (YOLOv3 is a multiscale detector, whereas for our application, objects are only present at a single scale). Finally, we modified the training loss function to suppress false positives and account for the large amounts of empty space in the training images. We achieved this by increasing the weighting (from 0.5 to 2) given to the no-object component of the multipart loss function described in Redmon et al. (2016). For training, we used transfer learning, again using the pretrained general purpose YOLO object detector as a starting point with initial weights created by training on the COCO dataset (Redmon & Farhadi, 2018). During training, we first froze all but the final 7 layers of the network and trained for 25 epochs, using the Adam optimiser and a learning rate of 10^{-4}. We then unfroze all layers, reduced the learning rate to $10^{-6,}$ and trained for a further 20 epochs. In total, training took 34 hours on a NVIDIA Quadro GP100 GPU. Other parameters of the algorithm, the detection threshold and bounding box overlap (non-maximum suppression) threshold, were selected based on minimising the difference between the automated count of the training images and the expert count. All code is available from http://dx.doi.org/10.5281/zenodo.2562058.

For the final stage, we counted 1,000 survey images selected at random, but excluding the 500 training images. Counting the test images took 2 hr using the same GPU as for the training.

2.4 | Expert count

The full set of images was counted, from 28 January to 28 February 2018, by a single expert counter (DJL) using Adobe CS6 as a viewing program operating on a Windows 10 operating system. Each JPEG was initially open in 'Fit' mode before enlarging to 50% or greater and counted using a left to right, top to bottom scanning pattern. Counting was conducting by clicking on and marking each wildebeest. The number of animals in each image was counted twice—first as running tally during marking and second as a recount of the marks within the image. While there remains the potential for bias in the expert count, we take this count to be the gold standard. Hence, our results are a comparison between the two novel methods employed and a count by a single experienced expert, which could in principle deviate from the unknown true count.

3 | RESULTS

We compared the accuracy of the methods by calculating the deviation of each method from the single expert count which we assume to be the true number of wildebeest in each image. For both the citizen science count and the YOLO count, we assess the

TABLE 1 Summary of key comparison statistics. Error rates are calculated across 1,000 survey images containing 19,802 wildebeest (Zooniverse) and 20,489 wildebeest (YOLO)

Method	Mean abs. error	RMS error	Percent overcount
Mean Zooniverse	2.8	8.11	−11.44%
Median Zooniverse	1.83	5.85	−8.85%
Filtered Zooniverse	1.58	4.05	0.55%
YOLO	1.70	5.94	0.69%

accuracy across 1,000 sample survey images. For the citizen science data, this means that empty images that were not uploaded to Zooniverse are included when assessing accuracy. In effect by comparing the methods in this way, we are assuming that completely empty images will be processed by citizen scientists with perfect accuracy.

To aggregate the Zooniverse data, we adopted three approaches. We took the mean of the 15 counts, the median of the 15 counts or removed the 5 lowest counts for each tile and took the mean of the remaining 10. For the last metric, the filtered mean, we determined which outliers to remove by minimising the root mean squared (RMS) error on the test dataset that is we removed five outliers in total but taking different numbers of highest or lowest counts, compared each combination to the expert counts and found that the optimal filtering was to remove the five lowest counts. We stress that the optimal filtering was determined on the same images used to assess the method hence there is no division of train and test images.

The per image error rates and total counts from the 1,000 images are shown in Table 1 and Figure 2. From these results, we see that all methods result in low per image error rate. However, while the average Zooniverse counts and the deep learning algorithm have similar RMS errors, there is a clear discrepancy when examining the total counts summed over all 1,000 images. The Zooniverse volunteers showed a systematic tendency to undercount the images; hence, there was approximately an 11% and 9% undercount for the total dataset for both the mean and the median. The YOLO algorithm did not show any systematic bias and although on average miscounted 1.7 wildebeest per image, its total was highly accurate, recording 20,631 wildebeest when the expert counted 20,489. Only the filtered mean, which averaged over the highest 10 volunteer counts, is comparable to the YOLO count in this respect.

We countered the systematic bias observed in the Zooniverse data by filtering the lowest 5 estimates and taking the mean of the remaining 10 to obtain a highly accurate count. In Figure 3, we show the cumulative counts across the 1,000 images for the aggregated Zooniverse data and the highest and lowest counts for each image. From this figure, it is clear that the expert count is not at the centre of the distribution of Zooniverse counters. Instead, a more accurate estimate is obtained by taking the mean of the 10 highest counts. It should be noted that both the mean and the median display this systematic bias; therefore, it is not simply due to the mean being a less robust estimator (Galton, 1907) but instead reveals a tendency for all Zooniverse counters, on average, to undercount. Whether this bias

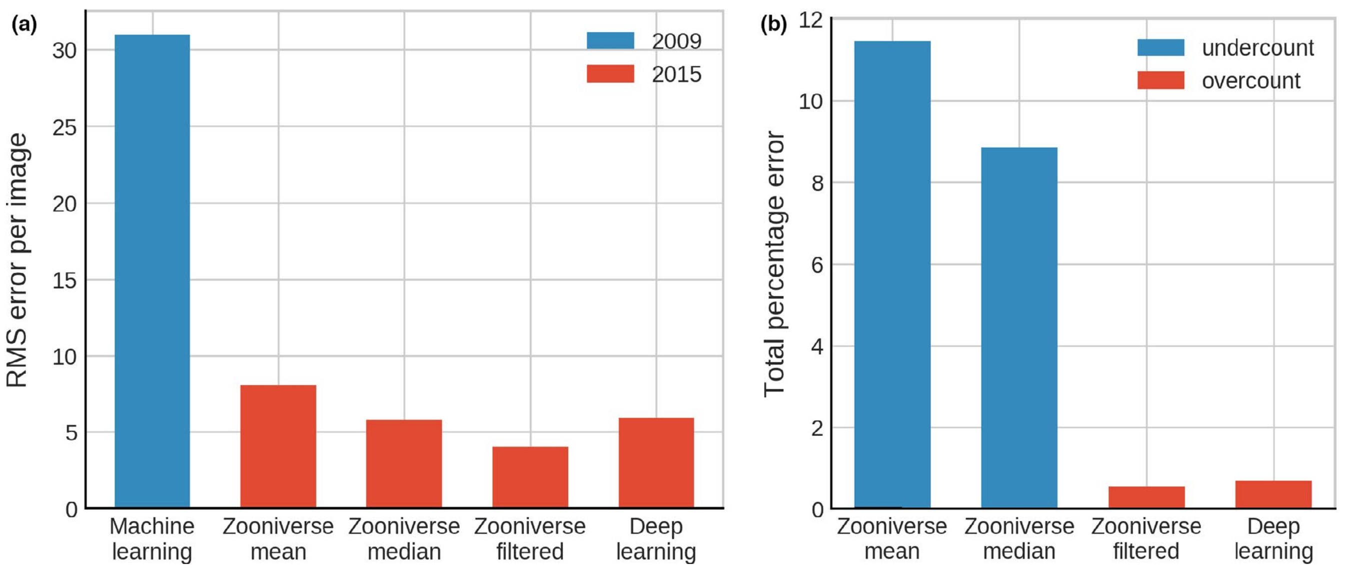

FIGURE 2 Error rates for the different methods calculated against the assumed true number of wildebeest per image as determined by the expert counter. (a) Root mean squared error calculated over 1,000 test images counted using the different methods. For comparison, we include results from (Torney et al., 2016), where handcrafted rotation invariant features were extracted and classified with a decision tree. It should be noted that the image resolution was lower in 2009 (4,288 × 2,848) and therefore a greater error should be expected. (b) The total percentage under or over count for 1,000 survey images for the Zooniverse data (aggregated using three different methods) and the deep learning algorithm

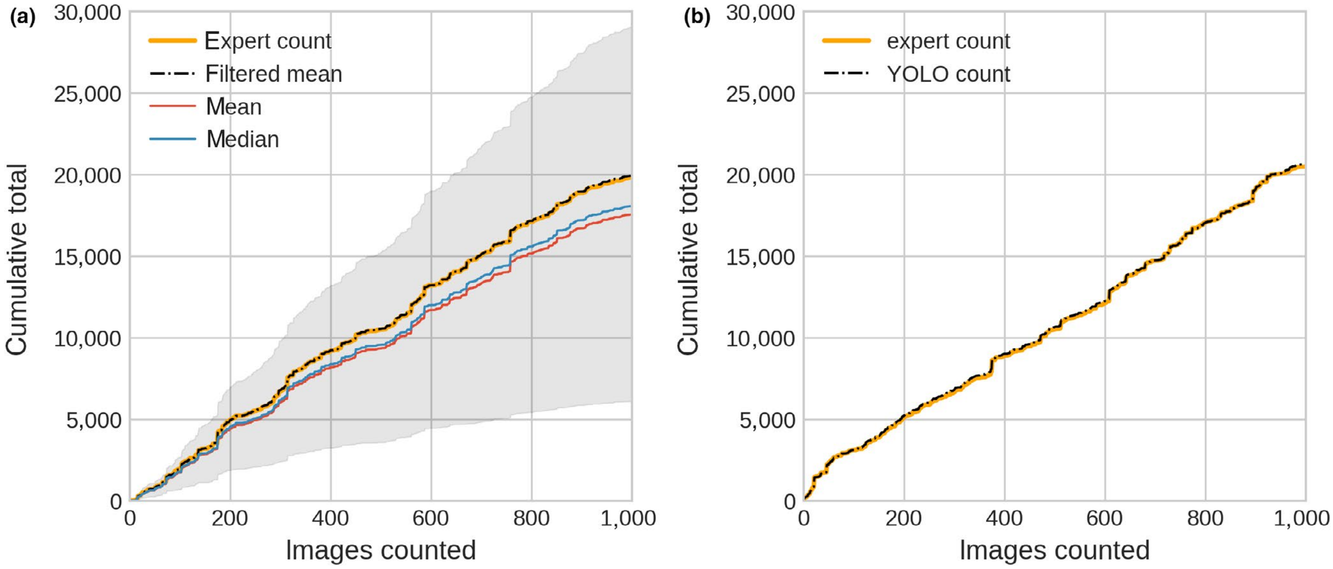

FIGURE 3 Cumulative image counts. (a) Mean, median and filtered mean for the Zooniverse count data compared to the expert count. The shaded region indicates the cumulative count that would be recorded if the highest or lowest counts for each tile of each image were used. (b) The YOLO count compared with the expert count

is persistent or predictable can only be revealed by repeated citizen science counts of the survey and comparison to expert counts.

4 | DISCUSSION

From our results, we see that both citizen science and deep learning methods are capable of producing highly accurate image counts. Counting the wildebeest within the survey images is a difficult and time-consuming task. When collecting the census images, there are multiple trade-offs between aircraft height, flight speed and camera parameters (ISO, exposure, etc.) that have to be balanced, with the result that image quality is often inconsistent. While wildebeest are often clear and unambiguous (see Figure 1c), in many cases, a subjective judgement has to be made based on the balance of probabilities that is what other animals are in the vicinity or what landscape features are present. In this context, we should not expect perfect agreement between our methods but estimates within 1% of the total can be considered as highly accurate.

For the citizen science counts, we observe a systematic bias in the errors the counters made. These results suggest that for a volunteer scientist the probability to miss a wildebeest is greater than the probability of incorrectly identifying another object or animal as a wildebeest. This is in line with prior expectations; given some guidance on identifying wildebeest (as was available on the Zooniverse project page), false positives should be minimal. However, eliminating false negatives requires substantial focus, and it is likely that concentration will wane over time, or volunteers will become distracted. Highly populated images have to be meticulously annotated while, equally, seemingly empty images have to be carefully scanned.

We found that it was possible to correct for this bias by removing the lowest five estimates. While this gives highly accurate total counts and low per image error rates, there is no guarantee that the approach is transferable and how to appropriately filter the data may be affected by the wording of the guidelines, the image resolution and sizes used, or the set of volunteers that participate in the project. Other more sophisticated approaches to processing citizen science data have been proposed (Swanson et al., 2016); however, given the range of counts provided by the volunteers and the large errors we observe in the baseline metrics (*c*. 11% and *c*. 9% undercount for the mean and median, respectively), there will need to be a rigorous process of validation before a citizen science count could be used as the sole basis for a population estimate. Note that if we simply took the minimum or maximum count per image the total count could have been either half or almost double the true count. While suppression of these types of outliers is fundamentally part of the ethos of citizen science, this does illustrate the broad range of responses from volunteers.

Considering the deep learning algorithm, we find that with minor modification and bespoke training, the object detection network proposed by Redmon and Farhadi (2018) is able to rapidly count 1,000 images and come up with a total that is within 1% of an expert count. As other authors have shown, DCNNs are able to process wildlife images for classification tasks (Chen, Han, He, Kays, & Forrester, 2014; Norouzzadeh et al., 2018; Villa, Salazar, & Vargas, 2017) and also detect and localise animals (Maire, Alvarez, & Hodgson, 2015; Schneider, Taylor, & Kremer, 2018). YOLO has the advantage of being a single-pass object detector that is fast and accurate. The 1,000 images can be processed in under 2 hours, meaning every future census could be counted within 24 hours. Hence, a process that currently takes 3–6 weeks, involving 3–4 wildlife professionals and countless cups of tea, can potentially be replaced with an automated system that runs overnight.

Image classification using pretrained DCNNs with state-of-the-art architectures (He, Zhang, Ren, & Sun, 2016; Simonyan & Zisserman, 2014) can be achieved with a few lines of code using open source libraries such as Keras (Chollet et al., 2015), while object detection algorithms are increasingly being integrated into libraries such as TensorFlow (Abadi et al., 2016). Currently, the greatest challenge for implementing these algorithms for bespoke applications is obtaining sufficiently large training datasets. In this regard, citizen scientists have a clear role to play. While we have shown that the trained algorithm achieves high accuracy levels, it should be noted that the algorithm employed the crowd-sourced data to create the training sets. Hence, both methods should be viewed as complementary approaches with citizen science data forming the foundation for automated algorithms (Rey et al., 2017).

Our results show that deep learning algorithms are now at a state where they can legitimately replace manual counters and remove a large burden from conservation organisations. The further great advantage of automated image processing is that it will allow us to leverage emerging image collection technologies, such as unmanned aerial vehicles, satellite platforms or fixed camera traps; coupling these advances in image collection tools with automated processing will greatly increase the accuracy of population estimates. As we move towards fully automated wildlife counts, it only remains to ensure the availability of sufficient training data that is representative of all potential survey images. This can be achieved by combining state-of-the-art deep learning methods with validated crowd-sourced training data.

An automated wildebeest image count will not only be a significant benefit for monitoring this specific population but also provides a transferable methodology that can be deployed for any population monitoring that currently includes manual image counts.

ACKNOWLEDGEMENTS

J.G.C.H. acknowledges support from the British Ecological Society large grant scheme, the Friedkin Foundation, and the European Union Horizon 2020 grant No 641918. C.J.T. acknowledges support from a James S. McDonnell Foundation Studying Complex Systems Scholar Award. We thank Anthony Dell for comments that improved the manuscript; we gratefully acknowledge the effort of the Zooniverse volunteers who counted wildebeest and thank Alexandra Swanson for assistance with data extraction code and comments on the manuscript. This publication uses data generated via the Zooniverse.org platform, development of which is funded by generous support, including a Global Impact Award from Google, and by a grant from the Alfred P. Sloan Foundation.

AUTHOR CONTRIBUTIONS

C.J.T., J.C.G.H. and D.J.L.-J. designed the study. D.J.L.-J., D.C.M., H.T.M., M.M., E.M.K. and J.G.C.H. collected the data. C.J.T., D.J.L.-J., M.C. and J.G.C.H. analysed the data. C.J.T. wrote the manuscript with input from all authors.

DATA ACCESSIBILITY

Raw count data are available from Enlighten: Research Data (https://doi.org//10.5525/gla.researchdata.732). Source code is available from http://dx.doi.orgl10.5281/zenodo.2562058.

ORCID

Colin J. Torney https://orcid.org/0000-0003-1673-7835

REFERENCES

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., … Isard, M. (2016). Tensorflow: A system for large-scale machine learning, in: OSDI, pp. 265–283.

Bajzak, D., & Piatt, J. F. (1990). Computer-aided procedure for counting waterfowl on aerial photographs. *Wildlife Society Bulletin*, *18*, 125–129.

Bruijning, M., Visser, M. D., Hallmann, C. A., & Jongejans, E. (2018). trackdem: Automated particle tracking to obtain population counts and size distributions from videos in r. *Methods in Ecology and Evolution*, *9*, 965–973. https://doi.org/10.1111/2041-210X.12975

Campbell, K., & Borner, M. (1995). *Population trends and distribution of Serengeti herbivores: Implications for management* (pp. 117–145). Serengeti II: Dynamics, management, and conservation of an ecosystem.

Chabot, D., Dillon, C., & Francis, C. (2018). An approach for using off-the-shelf object-based image analysis software to detect and count birds in large volumes of aerial imagery. *Avian Conservation and Ecology*, *13*, 227–251.

Chen, G., Han, T. X., He, Z., Kays, R., & Forrester, T. (2014). Deep convolutional neural network based species recognition for wild animal monitoring. *Image Processing (ICIP)* (pp. 858–862). 2014 IEEE International Conference on, IEEE.

Chollet, F., et al., (2015). Keras.

Christin, S., Hervet, E., & Lecomte, N. (2018). Applications for deep learning in ecology. bioRxiv, 334854.

Condorcet, M., (1976). Essay on the application of mathematics to the theory of decision-making. Reprinted in Condorcet: Selected Writings, Keith Michael Baker, ed 33.

Dell, A. I., Bender, J. A., Branson, K., Couzin, I. D., de Polavieja, G. G., Noldus, L. P., … Wikelski, M. (2014). Automated image-based tracking and its application in ecology. *Trends in Ecology and Evolution*, *29*, 417–428. https://doi.org/10.1016/j.tree.2014.05.004

Dill, K. A., & MacCallum, J. L. (2012). The protein-folding problem, 50 years on. *Science*, *338*, 1042–1046. https://doi.org/10.1126/science.1219021

Dobson, A. P., Borner, M., Sinclair, A. R. E., Hudson, P. J., Anderson, T. M., Bigurube, G., … Wolanski, E. (2010). Road will ruin Serengeti. *Nature*, *467*, 272–273. https://doi.org/10.1038/467272a

Ellwood, E. R., Crimmins, T. M., & Miller-Rushing, A. J. (2017). Citizen science and conservation: Recommendations for a rapidly moving field. *Biological Conservation*, *208*, 1–4

Estes, R. (2014). *The Gnu's world: Serengeti wildebeest ecology and life history*. Berkeley, CA: Univ of California Press.

Forrester, T. D., Baker, M., Costello, R., Kays, R., Parsons, A. W., & McShea, W. J. (2017). Creating advocates for mammal conservation through citizen science. *Biological Conservation*, *208*, 98–105. https://doi.org/10.1016/j.biocon.2016.06.025

Galton, F. (1907). Vox populi (the wisdom of crowds). *Nature*, *75*, 450–451. https://doi.org/10.1038/075450a0

Girshick, R., (2015). Fast r-cnn, in: Computer Vision (ICCV), 2015 IEEE International Conference on, IEEE. pp. 1440–1448.

Harris, G., Thirgood, S., Hopcraft, J. G. C., Cromsigt, J. P., & Berger, J. (2009). Global decline in aggregated migrations of large terrestrial mammals. *Endangered Species Research*, *7*, 55–76. https://doi.org/10.3354/esr00173

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778).

Holdo, R. M., Fryxell, J. M., Sinclair, A. R. E., Dobson, A., & Holt, R. D. (2011). Predicted impact of barriers to migration on the Serengeti wildebeest population. *PLoS ONE*, *6*, e16370. https://doi.org/10.1371/journal.pone.0016370

Holdo, R. M., Holt, R. D., Sinclair, A. R. E., Godley, B. J., & Thirgood, S. (2011). Migration impacts on communities and ecosystems: Empirical evidence and theoretical insights. In E. J. Milner-Gulland, J. M. Fryxell, & A. R. E. Sinclair (Eds.), *Animal Migration* (pp. 130–143). Oxford, UK: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199568994.001.0001

Jolly, G. (1969). Sampling methods for aerial censuses of wildlife populations. *East African Agricultural and Forestry Journal*, *34*, 46–49. https://doi.org/10.1080/00128325.1969.11662347

Kao, A. B., Berdahl, A. M., Hartnett, A. T., Lutz, M. J., Bak-Coleman, J. B., Ioannou, C. C., … Couzin, I. D. (2018). Counteracting estimation bias and social influence to improve the wisdom of crowds. *Journal of The Royal Society Interface*, *15*, 20180130. https://doi.org/10.1098/rsif.2018.0130

Kosmala, M., Wiggins, A., Swanson, A., & Simmons, B. (2016). Assessing data quality in citizen science. *Frontiers in Ecology and the Environment*, *14*, 551–560. https://doi.org/10.1002/fee.1436

Laliberte, A. S., & Ripple, W. J. (2003). Automated wildlife counts from remotely sensed imagery. *Wildlife Society Bulletin*, *31*, 362–371.

Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., … Zitnick, C. L. (2014). Microsoft COCO: Common objects in context. In D. Fleet, T. Pajdla, B. Schiele, & T. Tuytelaars (Eds.), *European conference on computer vision* (pp. 740–755). Cham: Springer.

Lintott, C. J., Schawinski, K., Slosar, A., Land, K., Bamford, S., Thomas, D., … Andreescu, D. (2008). Galaxy zoo: Morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey. *Monthly Notices of the Royal Astronomical Society 389*, 1179–1189. https://doi.org/10.1111/j.1365-2966.2008.13689.x

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). SSD: Single shot multibox detector. In B. Leibe, J. Matas, N. Sebe, & M. Welling (Eds.), *European conference on computer vision* (pp. 21–37). Cham: Springer.

Lovejoy, T. E., Sallaberry, M., Senner, S. E., & Tarak, A. (1987). Conservation strategy for migratory species. *American Scientist*, *75*, 19–26.

Mac Aodha, O., Gibb, R., Barlow, K. E., Browning, E., Firman, M., Freeman, R., … Newson, S. E. (2018). Bat detective—deep learning tools for bat acoustic signal detection. *PLoS Computational Biology 14*, e1005995. https://doi.org/10.1371/journal.pcbi.1005995

Maire, F., Alvarez, L. M., & Hodgson, A. (2015). Automating marine mammal detection in aerial images captured during wildlife surveys: A deep learning approach, In B. Pfahringer & J. Renz (Eds.), *Australasian Joint Conference on Artificial Intelligence* (pp. 379–385). Cham: Springer.

McNaughton, S. J. (1985). Ecology of a grazing ecosystem: The Serengeti. *Ecological Monographs*, *55*, 259–294. https://doi.org/10.2307/1942578

McNeill, S., Barton, K., Lyver, P., & Pairman, D. (2011). Semi-automated penguin counting from digital aerial photographs. *Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International, IEEE* (pp. 4312–4315). Vancouver, BC: IEEE. https://doi.org/10.1109/IGARSS.2011.6050185

Mduma, S. A., Sinclair, A., & Hilborn, R. (1999). Food regulates the Serengeti wildebeest: A 40-year record. *Journal of Animal Ecology*, *68*, 1101–1122. https://doi.org/10.1046/j.1365-2656.1999.00352.x

Norouzzadeh, M. S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M. S., Packer, C., & Clune, J. (2018). Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. Proceedings of the National Academy of Sciences, 201719367.

Norton-Griffiths, M. (1973). Counting the Serengeti migratory wildebeest using two-stage sampling. *African Journal of Ecology*, *11*, 135–149. https://doi.org/10.1111/j.1365-2028.1973.tb00079.x

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779–788). Las Vegas, NV: IEEE.

Redmon, J., & Farhadi, A. (2017). YOLO9000: Better, Faster, Stronger, 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 6517–6525), Honolulu, Hawaii: IEEE. https://doi.org/10.1109/cvpr.2017.690

Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767.

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, & R. Garnett (Eds.), *Advances in neural information processing systems* (pp. 91–99). Montreal, Canada: Curran Associates, Inc.

Rey, N., Volpi, M., Joost, S., & Tuia, D. (2017). Detecting animals in African Savanna with UAVs and the crowds. *Remote Sensing of Environment*, *200*, 341–351. https://doi.org/10.1016/j.rse.2017.08.026

Sauermann, H., & Franzoni, C. (2015). Crowd science user contribution patterns and their implications. *Proceedings of the National Academy of Sciences*, *112*, 679–684. https://doi.org/10.1073/pnas.1408907112

Schneider, S., Taylor, G. W., & Kremer, S. C. (2018). Deep learning object detection methods for ecological camera trap data. arXiv preprint arXiv:1803.10842.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

Simpson, R., Page, K. R., & De Roure, D. (2014). Zooniverse: Observing the world's largest citizen science platform. In Chin-Wan Chung (Ed.), *Proceedings of the 23rd international conference on world wide web, ACM* (pp. 1049–1054). New York, NY: ACM.

Sinclair, A. R. E., Metzger, K. L., Mduma, S. A. R., & Fryxell, J. M. (2015). *Serengeti IV: Sustaining Biodiversity in a Coupled Human-Natural System*. Chicago: University of Chicago Press. https://doi.org/10.7208/chicago/9780226196336.001.0001

Singh, N. J., & Milner-Gulland, E. J. (2011). Conserving a moving target: Planning protection for a migratory species as its distribution changes. *Journal of Applied Ecology*, *48*, 35–46. https://doi.org/10.1111/j.1365-2664.2010.01905.x

Subalusky, A. L., Dutton, C. L., Rosi, E. J., & Post, D. M. (2017). Annual mass drownings of the Serengeti wildebeest migration influence nutrient cycling and storage in the Mara river. *Proceedings of the National Academy of Sciences of the United States of America*, *114*, 7647–7652. https://doi.org/10.1073/pnas.1614778114

Surowiecki, J. (2005). *The wisdom of crowds*. New York, NY: Anchor.

Swanson, A., Kosmala, M., Lintott, C., & Packer, C. (2016). A generalized approach for producing, quantifying, and validating citizen science data from wildlife images. *Conservation Biology*, *30*, 520–531. https://doi.org/10.1111/cobi.12695

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., … Rabinovich, A. (2015). Going deeper with convolutions. *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1–9). Boston, MA: IEEE.

Thirgood, S., Mosser, A., Tham, S., Hopcraft, G., Mwangomo, E., Mlengeya, T., … Borner, M. (2004). Can parks protect migratory ungulates? The case of the Serengeti wildebeest. *Animal Conservation*, *7*, 113–120. https://doi.org/10.1017/S1367943004001404

Torney, C. J., Dobson, A. P., Borner, F., Lloyd-Jones, D. J., Moyer, D., Maliti, H. T., … Hopcraft, J. G. C. (2016). Assessing rotation-invariant

feature classification for automated wildebeest population counts. *PLoS ONE 11*, e0156342. https://doi.org/10.1371/journal.pone.0156342
Turchin, P. (1999). Population regulation: A synthetic view. *Oikos*, *84*, 153–159. https://doi.org/10.2307/3546876
Valletta, J. J., Torney, C., Kings, M., Thornton, A., & Madden, J. (2017). Applications of machine learning in animal behaviour studies. *Animal Behaviour*, *124*, 203–220. https://doi.org/10.1016/j.anbehav.2016.12.005
Villa, A. G., Salazar, A., & Vargas, F. (2017). Towards automatic wild animal monitoring: Identification of animal species in camera-trap images using very deep convolutional neural networks. *Ecological Informatics*, *41*, 24–32. https://doi.org/10.1016/j.ecoinf.2017.07.004
Walters, C.J. (1986). *Adaptive management of renewable resources*. New York, NY: Macmillan Publishers Ltd.
Weinstein, B. G. (2018). A computer vision for animal ecology. *Journal of Animal Ecology*, *87*, 533–545. https://doi.org/10.1111/1365-2656.12780
Wilcove, D. S., & Wikelski, M. (2008). Going, going, gone: Is animal migration disappearing. *PLoS Biology*, *6*, e188. https://doi.org/10.1371/journal.pbio.0060188
Xue, Y., Wang, T., & Skidmore, A. K. (2017). Automatic counting of large mammals from very high resolution panchromatic satellite imagery. *Remote sensing*, *9*, 878. https://doi.org/10.3390/rs9090878
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*, e115989. https://doi.org/10.1371/journal.pone.0115989

SUPPORTING INFORMATION

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

How to cite this article: Torney CJ, Lloyd-Jones DJ, Chevallier M, et al. A comparison of deep learning and citizen science techniques for counting wildlife in aerial survey images. *Methods Ecol Evol*. 2019;10:779–787. https://doi.org/10.1111/2041-210X.13165