
There may be differences between this version and the published version. You are advised to consult the publisher’s version if you wish to cite from it.


This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.

[http://eprints.gla.ac.uk/180086/](http://eprints.gla.ac.uk/180086/)

Deposited on: 15 February 2019

Enlighten – Research publications by members of the University of Glasgow
[http://eprints.gla.ac.uk](http://eprints.gla.ac.uk)
A comparison of deep learning and citizen science techniques for counting wildlife in aerial survey images

Colin J. Torney\textsuperscript{a,}, David J. Lloyd-Jones\textsuperscript{b,\ast}, Mark Chevallier\textsuperscript{a}, David C. Moyer\textsuperscript{c}, Honori T. Maliti\textsuperscript{d}, Machoke Mwita\textsuperscript{d}, Edward M. Kohi\textsuperscript{d}, J. Grant C. Hopcraft\textsuperscript{e}

\textsuperscript{a}School of Mathematics and Statistics, University of Glasgow, Glasgow G12 8SQ, UK
\textsuperscript{b}FitzPatrick Institute of African Ornithology, DST-NRF Centre of Excellence, University of Cape Town, Rondebosch 7701, South Africa
\textsuperscript{c}Integrated Research Center, The Field Museum of Natural History, 1400 S. Lake Shore Drive, Chicago, United States of America
\textsuperscript{d}Tanzania Wildlife Research Institute, P. O. Box 661, Arusha, Tanzania
\textsuperscript{e}Institute of Biodiversity, Animal Health and Comparative Medicine, University of Glasgow, Glasgow G12 8QQ, UK

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, 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 (\textit{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.

\ast authors contributed equally
1. Introduction

Estimating the abundance of animal species is essential for ecologists, conservationists, and wildlife managers worldwide. 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 populations (Turchin [1999]), a means to determine their vital rates (Mduma et al. [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 et al. [2009], Singh and Milner-Gulland [2011]). Estimating population sizes of migratory...
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.
species is a vital, but logistically challenging task. Localized environmental disturbances in large geographic 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 et al., 1987; Wilcove and Wikelski, 2008).

The Serengeti National Park in Tanzania is known for the iconic migration of approximately 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 et al., 2011b; McNaughton, 1985; Subalusky et al., 2017), as well as providing vital ecosystem services to human communities around the park (Sinclair et al., 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 et al., 2011a). The long-term population trend of the wildebeest (see Fig. 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 and Borner, 1995; Norton-Giffiths, 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 georeferenced 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 approximately 3-6 weeks (Torney et al., 2016). Automating this aspect of the survey would have two major advantages. Firstly, it would remove a bottleneck in running the survey. Beyond the actual counting time, there is often a considerable delay in scheduling the counting process as it involves multiple wildlife professionals to undertake. Secondly, removing this time-consuming job would relieve a significant burden on the organizations 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 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, 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 to astronomy, and appears to be growing as a tool for ecologists and conservationists, where it has the major advantage of not only performing scientific analysis of data but also engaging the public with wildlife conservation. However, despite the growth in the use and awareness of citizen science approaches, there is still some skepticism about the reliability of unpaid and often anonymous volunteers (see 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 and have been applied previously to the task of counting aerial images of animals including the Serengeti wildebeest population. While attempts to automate the classification and/or localisation of objects within images have been on-going 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, 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. 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 specialized object detection networks have been developed that either use a two stage process of proposing regions then classifying them, or a single pass through the network to predict object classes and their coordinates. In this work we evaluate the performance of the single-pass DCNN architecture proposed in.
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 1000 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 (4288x2848) and therefore a greater error should be expected. (b) The total percentage under or over count for 1000 survey images for the zooniverse data (aggregated using 3 different methods) and the deep learning algorithm.

2. Methods

2.1. Aerial surveys

The 2015 Serengeti wildebeest count was conducted between the 23rd April and 2nd May over the eastern and southern plains of Serengeti National Park, Ngorongoro Conservation Area, Loliordo 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 35mm 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 seconds.

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 2040 km.

During the count, flight target altitude along transects was 700ft (213m) 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 7360x4912 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 the 10th of May and 31st of 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 and 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 864x864 sub-images and then passed though a version of the YOLO DCNN using pre-trained weights from the COCO dataset provided by (Redmon and 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 864x864 training images).

With this training set, we next made several minor modifications to the YOLOv3 architecture. YOLOv3 employs 9 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 9 to 3, and replaced
the dimensions of the 3 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
pre-trained general purpose YOLO object detector as a starting point with initial weights created by training on the
COCO dataset (Redmon and 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 https://github.com/ctorney/deepWildCount.

For the final stage we counted 1000 survey images selected at random, but excluding the 500 training images.
Counting the test images took 2 hours using the same GPU as for the training.

2.4. Expert count

The full set of images was counted, from January 28 to February 29 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 were counted
twice — first as running tally during marking and secondly 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 accuracy across 1000 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
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.

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 minimizing the root mean squared (RMS) error on the test dataset, i.e. we removed 5 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 5 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 1000 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 1000 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 Fig. 3 we show the cumulative counts across the
<table>
<thead>
<tr>
<th>Method</th>
<th>Mean abs. error</th>
<th>RMS error</th>
<th>Percent overcount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean zooniverse</td>
<td>2.8</td>
<td>8.11</td>
<td>-11.44%</td>
</tr>
<tr>
<td>Median zooniverse</td>
<td>1.83</td>
<td>5.85</td>
<td>-8.85%</td>
</tr>
<tr>
<td>Filtered zooniverse</td>
<td>1.58</td>
<td>4.05</td>
<td>0.55%</td>
</tr>
<tr>
<td>YOLO</td>
<td>1.70</td>
<td>5.94</td>
<td>0.69%</td>
</tr>
</tbody>
</table>

Table 1: Summary of key comparison statistics. Error rates are calculated across 1000 survey images containing 19,802 wildebeest (zooniverse) and 20,489 wildebeest (YOLO).

1000 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 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 Fig 1c), in many cases a subjective judgment has to be made based on the balance of probabilities, i.e. 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’s 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 5 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 (~11% and ~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 1000 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 et al., 2014; Norouzzadeh et al., 2018; Villa et al., 2017) and also detect and localise animals (Maire et al., 2015; Schneider et al., 2018). YOLO has the advantage of being a single-pass object detector that is fast and accurate. The 1000 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 to 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 et al., 2016; Simonyan and 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 approaches.
An automated wildebeest image count will not only be a significant benefit for monitoring this specific population but provides a transferable methodology that can be deployed for any population monitoring that currently includes manual image counts.

**Author contributions** CJT, JCGH, DJLJ designed the study. DJLJ, DCM, HTM, MM, EMK, JGCH collected data. CJT, DJLJ, MC, JGCH analysed the data. CJT wrote the manuscript with input from all authors.

**Acknowledgements** JGCH acknowledges support from the British Ecological Society large grant scheme, the Friedfeld Foundation, and the European Union Horizon 2020 grant No 641918. CJT 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.

**Data Accessibility** Raw count data are available from Enlighten: Research Data [http://dx.doi.org/10.5525/gla.researchdata.732]. Source code is available from [http://dx.doi.org/10.5281/zenodo.2562058].

---

12