Toward Using Citizen Scientists to Drive Automated Ecological Object Detection in Aerial Imagery

Connor Bowley*, Marshall Mattingly III*, Andrew Barnas†, Susan Ellis-Felege†, Travis Desell*
Department of Computer Science*, Department of Biology†
University of North Dakota
Grand Forks, North Dakota 58202
Email: {connor.bowley, marshall.p.mattingly, andrew.barnas, susan.felege}@und.edu, tdesell@cs.und.edu

Abstract*—Automated object detection within imagery is challenging in the field of wildlife biology. Uncontrolled conditions, along with the relative size of target species to the more abundant background makes manual detection tedious and error-prone. In order to address these concerns, the Wildlife@Home project has been developed with a web portal to allow citizen scientists to inspect and catalog these images, which in turn provides training data for computer vision algorithms to automate the detection process. This work focuses on a project with over 65,000 UAS images from flights in the Hudson Bay area of Canada gathered in the years 2015 and 2016. This data set comprises over 3TB or raw imagery and also contains a further 2 million images from related ecological projects. Given the data scale, the person-hours that would be needed to manually inspect the data is extremely high. This work examines the efficacy of using citizen science data as inputs to convolutional neural networks (CNNs) used for object detection. Three CNNs were trained with expert observations, citizen scientist observations, and matched observations made by pairing citizen scientist observations of the same object and taking the intersection of the two observations. The expert, matched, and unmatched CNNs overestimated the number of lesser snow geese in the testing images by 88%, 150%, and 250%, respectively, which is less than current work using similar techniques on all visible (RGB) UAS imagery. These results show that the accuracy of the input data is more important than the quantity of the input data, as the unmatched citizen scientists observations are shown to be highly variable, but substantial in number, while the matched observations are much closer to the expert observations, though less in number. To increase the accuracy of the CNNs, it is proposed to use a feedback loop to ensure the CNN gets continually trained using extracted observations that it did poorly on during the testing phase.

I. Introduction

The Wildlife@Home project is a *citizen science* web portal that allows non-field scientists to aid in the cataloging of data from ecological video and images. There is currently over 100,000 hours of video captured by nest cameras and over 65,000 aerial images captured with the usage of an unmanned aerial system (UAS) from multiple sub-projects, the majority of which still need cataloging. For field-scientists, called *experts* throughout this paper, to manually categorize every minute of video and every pixel of imagery is infeasible, especially within a timely manner. Utilizing citizen scientists to crowd source this categorization dramatically increases the throughput of cataloging.

The primary concern with citizen scientists is proving their ability to produce good, consistent observations that closely match observations an expert would produce. In prior work, it was shown that individual citizen scientists produce highly variable bounding boxes around objects from the UAS imagery [1]. This variability can be reduced by matching two or more citizen sciences observations to the same object and extracting the intersection of the observations to produce a singular matched observation that is less variable and more closely matches expert observations on the same object.

This paper expands upon prior work by demonstrating the viability of using citizen science observations when compared against expert observations for providing training data to automated machine learning using convolutional neural networks (CNNs). Proving the efficacy of citizen science observations being used for CNN inputs increases confidence that citizen science can be used to facilitate field-scientists to process data and reach conclusions quicker. In the relatively small geographical area of the dataset used in this paper, there are over 65,000 individual UAS images totaling over 3 TB of raw data, with near that number expected to be produced each year for the next several years. Citizen scientists are therefore instrumental in the initial data processing, especially given the expected increase in the data gathering rate.

CNNs are commonly used for image classification, making them well suited for use in this work. CNNs transform an input image through a number of hidden layers (convolutional, max-pooling, and fully connected layers are the most common types) and finally predict to which class the input image belongs. The convolutional and fully connected layers contain weights that are trained to allow the network to accurately identify the images. Typically training is done using gradient descent, with additional techniques having been developed to aid training.

This paper is divided into several sections. Section II discusses similar projects in the field, especially ecological citizen science projects. Section III provides the full details on the collected images in the dataset. Section IV describes the methodology used in gathering and processing the citizen science observations, preprocessing and formatting the data for usage in the CNNs, and the structure and physical systems used for the running and analysis of the CNNs. Section V presents the results of the CNNs and provides analysis and context for the results. Section VI reaches conclusions about the results. Finally, Section VII touches on potential future works of the project to increase the correctness of the CNNs.

II. Related Work

Crowd sourcing has become more and more popular in recent years, with many projects successfully using citizen scientists to produce results. In the astronomical fields, GalaxyZoo [2], [3] allows citizen scientists to classify galaxies in images from the Sloan Digital Sky Survey [4] and PlanetHunters [5] allows citizen scientists to identify planet candidates from the NASA Kepler public release data. Citizen scientists have also been successfully leveraged in ecological projects. Snapshot Serengeti [6] allows citizen scientists to identify objects from camera traps in the Serengeti National Park.

Zooniverse [7] is a platform that allows for rapid development of citizen science projects, including GalaxyZoo and Snapshot Serengeti. At the start of 2014, there were over 20 projects in a wide variety of fields hosted by Zooniverse [7], the best performing of which tend to be well-established projects relating to astronomy with good community outreach [8]. Having a robust platform that allows for scientists to get the data in the hands of citizen scientists quickly and easily is paramount for the usefulness of the platform. The major differentiator between this work and Zooniverse is the granularity of input. Zooniverse allows citizen scientists to say what and how many of the objects are in the images, while this work also allows the where to be defined. To give accurate, automated inputs to a CNN, the where is extremely important.

There are a few projects using citizen science within avian ecology specifically. NestCams [9] is a Cornell project that has video cameras installed primarily in bird houses, opportunistically capturing a variety of cavity-nesting species. CamClickr [10] was used to catalog nesting behavior in over 600,000 images. CamClickr was even used during a university biology curriculum to teach students how to accurately identify objects in images while being aware of potential observer biases [11]. eBird [12] allows users to upload user-taken images of bird observations through handheld devices, providing spatio-temporal information about the bird distribution and abundance.

Automated object detection from ecological imagery has started seeing more use in recent years. A recent project used mosaic UAS imagery of white-tailed deer with both visible (RGB) and thermal infrared (TIR) spectra [13]. Supervised and un-supervised pixel-based detection, which is a more simplistic detection method, were both unsuccessful on both RGB and TIR imagery; however, object-based image analysis (OBIA) proved to be extremely successful on the TIR imagery, producing no false-positives while matching the 50% detection rate of manned aerial surveys [13]. OBIA on the RGB images, however, had an extreme number of false positives, with 1,946 deer detected in an image with only 4 actual deer present, as an example [13]. This shows that TIR can potentially dramatically increase the accuracy of automated ecological object detection in UAS imagery and highlights the difficulty of RGB analysis.

Another project used video recorded during UAS flights and used feature-based analysis using two features — color and shape — to detect birds in the video [14]. After manually selecting the input objects for feature-testing, the system was able to have omission (missed objects) and commission (false-positives) rates of less that 20% each [14]. This is an encouraging method that can potentially be used in conjunction with CNNs in future work.

III. Wildlife@Home Image Dataset

A major consideration for using UAS is that of safety. Manned aerial wildlife surveys are a leading causes of death for wildlife biologists, accounting for 66% of work related mortality from 1937-2000 [15]. Using a UAS to perform the flights and record images and video for observation in a safe environment has therefore seen an increased usage in ecological and biological projects [16]–[26]. It was shown that sampling bias is prevalent in manned aerial wildlife surveys, even when looking at large mammals over a large area, with the surveys finding 50% or less of the number of groups of the target species [27], [28].

A Trimble UX5[1] fixed wing UAS was flown at Wapsusk National Park in Manitoba, Canada in summer 2015 and summer 2016. Flights were conducted during the nesting season of lesser snow geese and common eiders and approximately a month later during the post-hatch period. Flights were conducted at 75m, 100m, and 120m above ground level along pre-defined transects. Images were recorded with a 16 megapixel Sony red, green, blue (RGB) flown in the nadir position with an 80% overlap between images.

For ease of presentation, the overlapping images were used to generate mosaics using Trimble Business Center[2] (version 3.51) and Pix4D[3] (version 3.2.23) for the 2015 and 2016 images, respectively. These mosaic images were then split into 1024x768 sub-images for presentation to the citizen scientists on Wildlife@Home. Each mosaic comprises several hundred sub-images.

In total, the combined set of flights produced over 65,000 raw UAS images totaling over 3 TB of raw imagery. These images combined into 36 distinct mosaics totaling over 50 GB of mosaic imagery that was sliced into 8,759 sub-images for presentation on Wildlife@Home.

There was a technical issue during the 2015 flights that incorrectly filtered the blue channel of the camera, producing images that are skewed from true RGB images such as the 2016 data. To alleviate this issue for the CNNs, the 2015 data was RGB-shifted to match the 2016 data. This shifting is further discussed in Sec. IV-A.

[1] http://uas.trimble.com/ux5
[2] http://www.trimble.com/Survey/trimble-business-center.aspx
[3] https://pix4d.com/

IV. Methodology

The methodology of the project is described in terms of: *(i)* the crowd sourcing web portal design, implementation, and data; *(ii)* preprocessing the raw data for usage in CNNs; *(iii)* how the data was selected and formatted for the CNNs; *(iv)* how the CNNs were structured and trained; and *(v)* how results were evaluated and quantified.

A. Citizen Science Data Gathering

In prior work, a web portal was created as part of the Wildlife@Home project to allow citizen scientists to examine the ecological UAS imagery collected [1]. The citizen scientists are presented images from an observation queue and are instructed to draw boxes around all objects that they can find in the image. The citizen scientists are given documentation to help them identify objects accurately, draw the bounding boxes around the object in such a way as to minimize negative space, and submit the information to the database.

The current web portal can be seen in Figure 1. At the top of the image is a status bar, which is currently showing how many images are left for review in the current mosaic, with the message being replaced by success or error messages upon submission of the observations. On the right side of the interface is the image being shown to the user who is able to double-click (or double-tap) to create a new observation which can be resized and moved around the interface to identify an object. The image can be zoomed in and panned around, with the current zoom level being shown in the bottom right and the current location within the image being shown by the scroll bars on the bottom and right.

On the left side of the interface is the internal image number with a balloon button that opens a post on the Wildlife@Home forum for discussion about the image and help buttons for species and the interface, which open popups with instructions on identifying different species or how to use the interface, respectively. Next are the object identification forms for each observation drawn in the image. Users are able to change the species, note whether or not they believe the species is on a nest, or delete the observation and corresponding box. Finally, users are able leave general comments about the image and submit their observations to the database, or submit that there is nothing here if no observations are made on the image.

There have been 63 unique citizen scientists who have made 6,852 observations. From the citizen scientist observations, 3,894 have been matched with at least one other citizen scientist to create matched observations using the 10-pixel corner point method from prior work [1]. There have been 5 unique experts who have made 2,775 observations, 1,200 of which are from a single expert who completed 12 mosaics.

To encourage citizen scientists to make observations, gamefication was implemented to give citizen scientists points for making observations, which unlock different badges on the Wildlife@Home forum and is used for a public leaderboard of image reviewers[4]. Citizen scientists are given 1-point for completing an image, whether or not any observations were made, 2-points for each observation made within the image, and an additional 5-points for each observation that is matched with another citizen scientist. This point system was designed to promote good results over simply submitting nothing to the database to earn 1-point at a time. In the future, more points will be given for observations that have been confirmed to be true, and potentially points may be deducted for observations that are confirmed to be false.

[4]https://csgrid.org/csg/top_image_reviewers.php

B. Correcting Blue-Shifted Imagery

There was a technical fault with the RGB camera used during collection of the 2015 UAS imagery that affected the blue channel. After reviewing several images between the 2015 and 2016 datasets, a simple multiplier on each channel was used to normalize the 2015 images to the 2016 images. Every pixel in the 2015 images had the red-, green-, and blue channels multiplied by $233.0/150.0$, $255.0/189.0$, and $236.0/190.0$, respectively, and floored to the nearest integer with a maximum of 255 on each channel. This appeared visually correct, as seen in Fig. 2, and proved to normalize the RGB spectrum well enough for use in the CNNs with both 2015 and 2016 datasets combined.

The alternative would be to leave the 2015 and 2016 datasets distinct and run CNNs for each dataset independently. Given the relatively small number of observations and the encouraging results of the normalized dataset, it was determined that the more numerous combined and normalized dataset with 2015 and 2016 images was better for CNN training than two distinct datasets.

C. Preprocessing the Data and Data Formats

The citizen scientists identified locations of white and blue phase lesser snow geese on images ranging in size from 1024x768 pixels to over 2000x3000 pixels. Typical CNNs have a fixed input size ranging from 28x28 pixels in the MNIST dataset [29] to a few hundred pixels on each dimension for the ImageNet dataset [30]. For datasets with some variation in sizes, such as ImageNet, it is common to randomly crop one or more images of a fixed size and run the cropped images through the CNN. Due to the nature of the data in this work, in which images are large and the foreground being searched for in the images are small, and the fact that counting the number of geese in an image is a goal, not just detection, randomly cropping images from an image that contains geese and labeling those images as foreground was not deemed acceptable.

To deal with the disparity in the sizes of data and what the CNN needs, two formats for the data were used. First, for both training and testing the CNN, 18x18 pixel sub-images were extracted from the larger images, a size chosen because objects in the imagery typically ranged from 14 — 18 pixels in each dimension. Images presented to users were from varying heights of 75m, 100m, and 125m. It was decided to group observations from these heights into a single dataset as the end observations are within a couple pixels of each other in

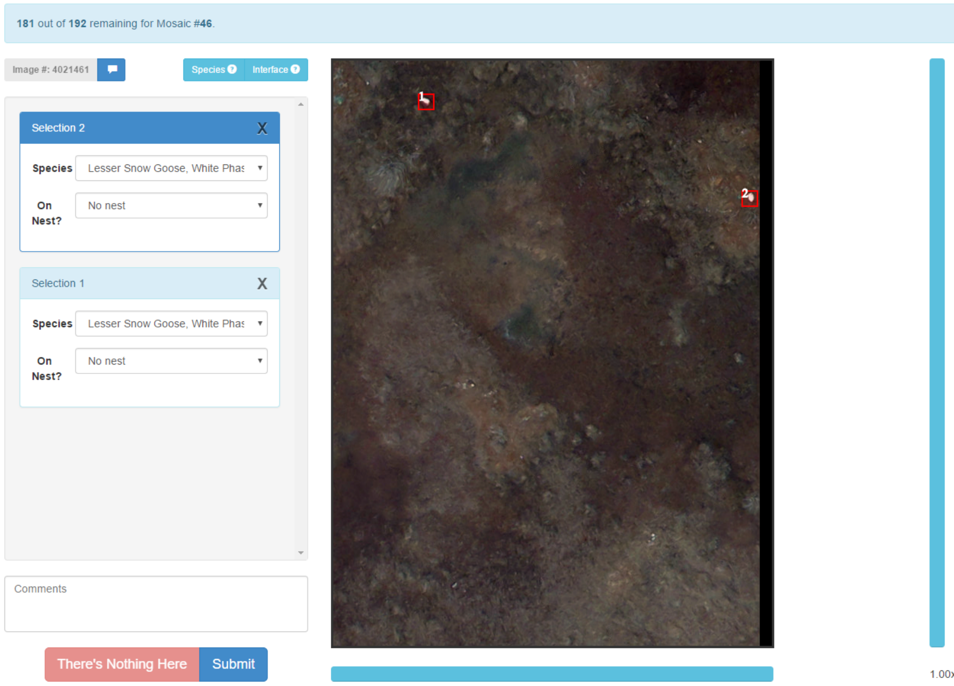
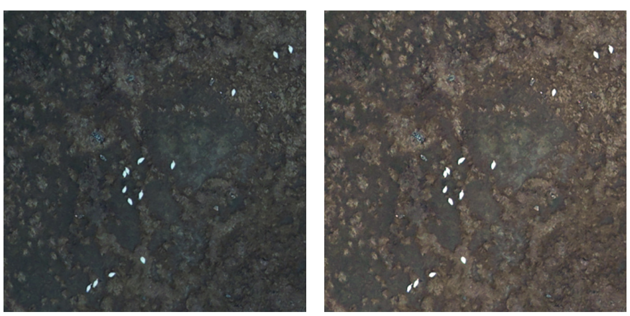


Fig. 1. The graphical user interface (GUI) for identifying objects in ecological imagery for the Wildlife@Home projects. This screenshot shows a UAS image with two white snow geese identified by the user on mosaic 46 with 181 out of 192 images remaining in the mosaic.



(a) Original image from 2015 with blue-shift error

(b) Same image from 2015 after the normalizing algorithm

Fig. 2. An example of the blue-shift error on a 2015 UAS image with the resultant image after RGB normalization to closely match the RGB spectrum of the 2016 UAS imagery. The white snow geese are actually white and the ground is correctly brown in the normalized image.

each dimension. These sub-images were stored in the IDX file format and will be referred to as training and testing IDXs. The other format used was PNG images of varying sizes taken straight from the mosaics and are referred to as mosaic split images (MSIs). The MSIs were used for testing purposes only.

For creating the IDX files for training and testing, two parts of data were considered separately. The two parts are foreground, which is any part of an image that contains a white or blue phase snow goose, and background, which is all other parts of the image. For extracting foreground, all boxes that were marked as either white phase or blue phase lesser snow geese were pulled from the images. For the background, random cropping was used to pull random background from the larger images. If one of the random crops overlapped at all with an area marked as a goose by any expert or citizen scientist, it was not included in the background set.

For testing a CNN over the MSIs, a striding method [31] was used to break the MSIs into smaller images (sub-images) that could be run through the CNN. The obvious downside to striding is that the amount of data that must be go through the CNN is substantially larger. During testing, however, more information is gained through this method because not only is the *what* predicted (as in what class the CNN labels a sub-image) but also the *where*. By reconstructing the MSI using the predictions on the sub-images, it can easily be seen where the geese are predicted to be in the image. A reconstruction of an MSI using the predictions by a CNN is referred to as a prediction image. Once the prediction of the MSI is created, blob counting techniques can be used to count the number of geese found by the CNN.

D. How the Data Was Chosen and Formatted

All imagery that had been labeled by both expert and citizen scientists were considered for training and testing. Approximately 20% of the MSIs that did not contain any geese and 20% of the MSIs that did were set aside for testing purposes.

From the remaining data, three separate datasets were made. These were an expert dataset containing only expert observations, a matched dataset containing only citizen scientist observations that were able to be matched, and an unmatched dataset containing all citizen scientist observations regardless

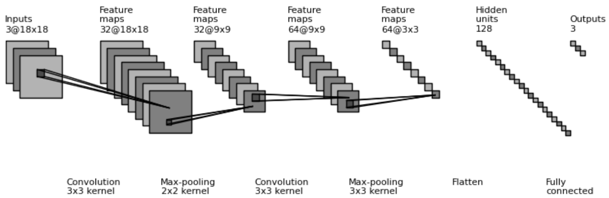


Fig. 3. Visualization of the CNN Architecture used in this work.

TABLE I
CNN ARCHITECTURE

Layer Type	Layer Dims	Filter / Pool Size	Stride	Filters	Padding
Input	18 x 18 x 3				
Convolutional	18 x 18 x 32	3	1	32	1
Max Pooling	9 x 9 x 32	2	2		
Convolutional	9 x 9 x 64	3	1	64	1
Max Pooling	3 x 3 x 64	3	3		
Fully Connected	1 x 1 x 128			128	
Fully Connected	1 x 1 x 3			3	

of whether or not they could be matched. The background images used in each of these datasets was identical.

Testing IDX files were created from the 20% of the data set aside for that purpose, and the MSIs themselves were used for testing. The IDXs created from this data contained only geese found by the experts, as that is considered the most true data and is what the citizen data is compared against. All of the CNNs were run against the same test data so they could be directly compared.

E. Training the CNNs

The CNN architecture used can be found in Figure 3 and Table I. It was trained on a Mac Pro using a 3.5 GHz 6-Core Intel Xeon E5 processor. Running over the MSIs used the processor and two AMD FirePro D700s concurrently. Nesterov Momentum with a momentum constant of 0.9 and L2 regularization with a λ of 0.05 were used to aid training. The learning rate started at $1x10^{-3}$ and was multiplied by 0.75 each epoch. Batch normalization was used after each convolutional layer and each fully connected layer. Batch normalization was placed before the activation function and used a minibatch size of 32, as found in the the original batch normalization paper [32]. The Leaky RELU activation function was used [33]. The weights for the convolutional and fully connected layers were randomly initialized based on a normal distribution with a mean of 0 and standard deviation of $\sqrt{2/n}$ with n being the number of inputs to the neuron. For batch normalization, there was one γ and β for each feature map and all γs were initialized to 1 and βs to 0. A separate CNN was trained on each dataset for 15 epochs and the epoch that had the best results on the training data was saved.

To deal with the disparity in the number of images in each class, undersampling of the background was used. There were 5 background images shown to the CNN for every 1 white phase snow goose image. This means that not every background image was shown every epoch. The background images shown were randomly selected each epoch.

Three CNNs were trained for this work, one on each dataset. These CNNs only identified white phase snow geese; blue phase snow geese were considered background.

F. Evaluation of Results

Once the CNNs were trained, the results were quantified in a few ways. The first way was the running the trained CNNs over an IDX file that contained novel testing images. The testing images that contained white phase snow geese were ones that were identified as such by the experts. This is the only evaluation method that does not use the prediction images over the MSIs.

The second of the quantification methods uses the prediction image to determine if a CNN "found" the observations the expert found. An observation is considered "found" by the CNN if 10% of the pixels within the bounding box drawn by the expert are considered the same species as identified by the expert. The reason the percentage is so low is that if a bounding box was unnecessarily large, the CNN could have marked it as the correct species but failed the "finding" process. Also, given the small percentage of pixels that are not background, it is unlikely the CNN "accidentally" marked part of the area in the box as the correct species. The percentage of background pixels that are misidentified as a goose is also recorded.

The final quantification method involves using a blob counter on the prediction image to count the number of blobs of each species. If the CNN is accurate, counting the blobs should give a rough estimate of the number of each species in the image. It should be noted that if two geese are located right next to each other in the image, it is likely that only one distinct blob will be created and counted. This is quite common as there are often nesting pairs of geese. Remedying this problem will be looked into for future work. The blob detector used was OpenCV's SimpleBlobDetector.

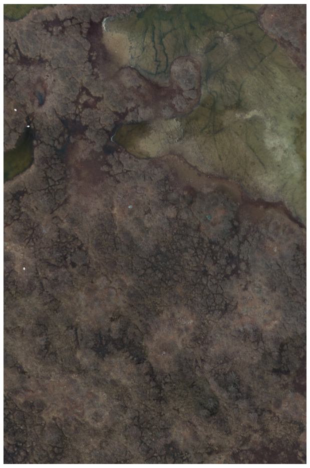
G. Data Statistics

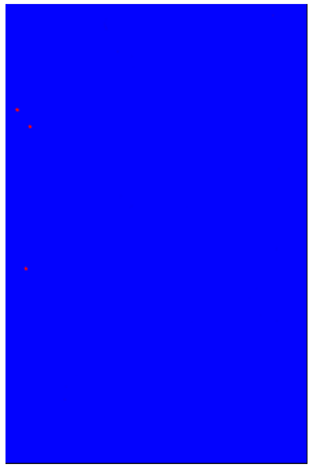
Table II shows the number of observation objects and background objects, each object being 18x18 pixels as described in Sec. IV-C, for all the datasets. As the CNNs were trained to identify white phase only, the blue phase observations are included in the background observations. This was done because white phase and blue phase snow geese share the same physical shape, with only the coloration being different. By adding the alternate phase observations to the background objects, the CNN is less likely to misidentify blue phase snow geese as white phase.

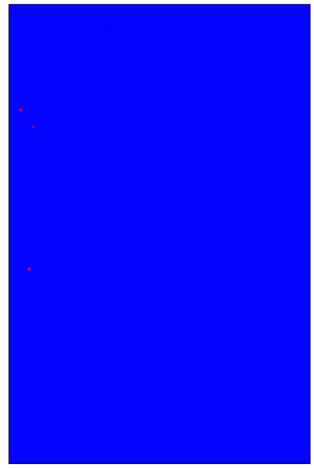
V. RESULTS

A. Training and Testing IDXs

The CNNs were training and tested using the datasets described in Sec. IV-G. These datasets were stored in the IDX file format described in Sec. IV-C.


(a) Original Image


(b) Prediction by expert CNN


(c) Prediction by matched CNN

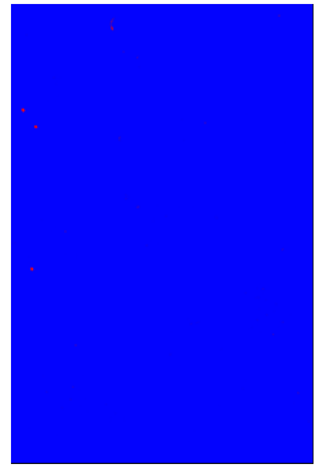

(d) Prediction by unmatched CNN

Fig. 4. Predictions on an image by three different CNNs. The expert trained CNN had less misclassified errors than the matched or unmatched CNNs.

TABLE II
DATA STATISTICS PER DATASET

Dataset	White Phase	Blue Phase	Background
Expert	1,731	503	78,040
Matched	5,344	405	78,040
Unmatched	6,264	1,043	78,040
Testing	401	129	18,278

TABLE III
IDX DATASET VS CNN TRAINING ACCURACY

	Accuracy		
CNN	White Phase	Background	Total
Expert	97.69%	99.81%	99.47%
Matched	98.65%	99.76%	99.58%
Unmatched	92.27%	99.14%	98.00%

TABLE IV
IDX DATASET VS CNN TESTING ACCURACY

	Accuracy		
CNN	White Phase	Background	Total
Expert	96.26%	99.91%	99.81%
Matched	91.52%	99.95%	99.76%
Unmatched	95.26%	99.75%	99.64%

As seen in Table III, the unmatched dataset had the lowest training percentages on the white phase and background for the training, which could indicate that the objects labeled as white phase in that dataset might have more variability to them than in the other datasets. Logically, this would make sense, as this dataset had observations from non-expert citizen users and no validation of their correctness was done. The highest overall accuracy on the training data came from the matched dataset with an overall training accuracy of 99.58%. The matched dataset also had the best accuracy on the white phase, while the expert dataset did best on the background.

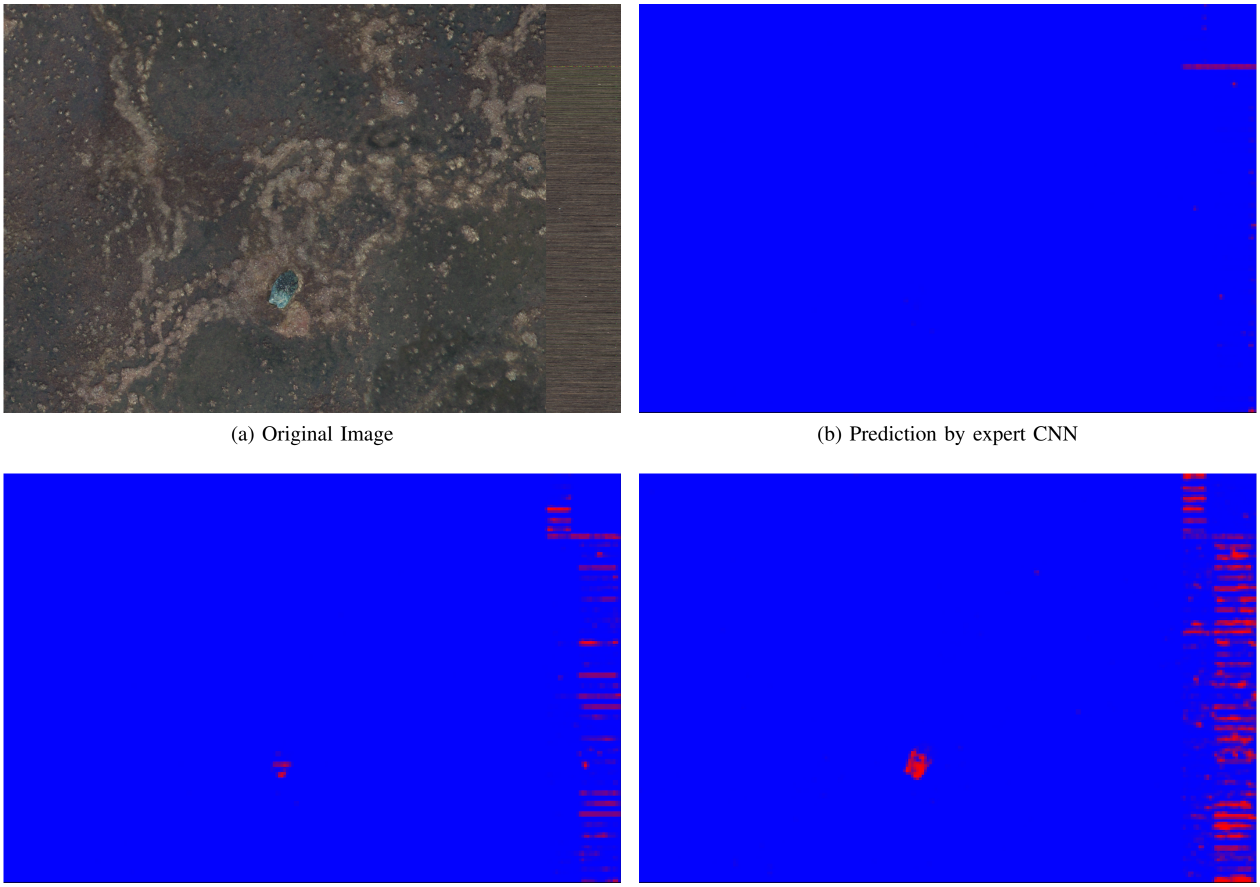
Of course, accuracy on training data is not nearly as important as accuracy on a novel set of test data. As seen in Table IV, the expert did the best overall, which would be expected. For the other networks, the unmatched one did better on identifying the white phase, but slightly worse on the background than the matched dataset. In this it should be noted that even though the network that used the matched data did considerably worse on the white phase and only marginally better on the background, it still had a better overall performance due to the much larger number of background images in the test set.

B. Testing MSIs

The IDX files used in Sec. V-A contain only a small portion of the total imagery. To prove the viability of each CNN, whole images are run through the network and a prediction image is created with the predicted classifications for every pixel of the original image. Using a blob counter, as described in Sec. IV-F, the number of observations made by the CNN can be compared against the known true number of observations made by experts on the same images.

1) Comparing Against User Observations: The unmatched CNN only missed 22 of the 401 white phase snow geese found by the experts, although it also had largest amount of misclassified background at 0.09%. The expert CNN did the best in terms of balancing both finding the white phase geese and not misclassifying background, only missing 26 geese and misclassifying 0.04% of the background, over twice as good as the unmatched. The matched dataset missed the most white phase at 41, and the amount of background it misclassified was at 0.05%, in between the expert and unmatched.

It must be noted that while the unmatched CNN missed the fewest geese, if one predicted that every pixel was white phase then no geese would be missed. That is why accurately



(a) Original Image

(b) Prediction by expert CNN

(c) Prediction by matched CNN

(d) Prediction by unmatched CNN

Fig. 5. Predictions on an image with artifacting. All the CNNs had some errors over the artifacted image with the expert CNN having the least errors. Also the expert CNN correctly identified the rock in the bottom center of the image as background while the matched and unmatched CNNs incorrectly identified parts of it as white phase.

TABLE V
COMPARISON OF CNN TO EXPERT OBSERVATIONS (WHITE ONLY)

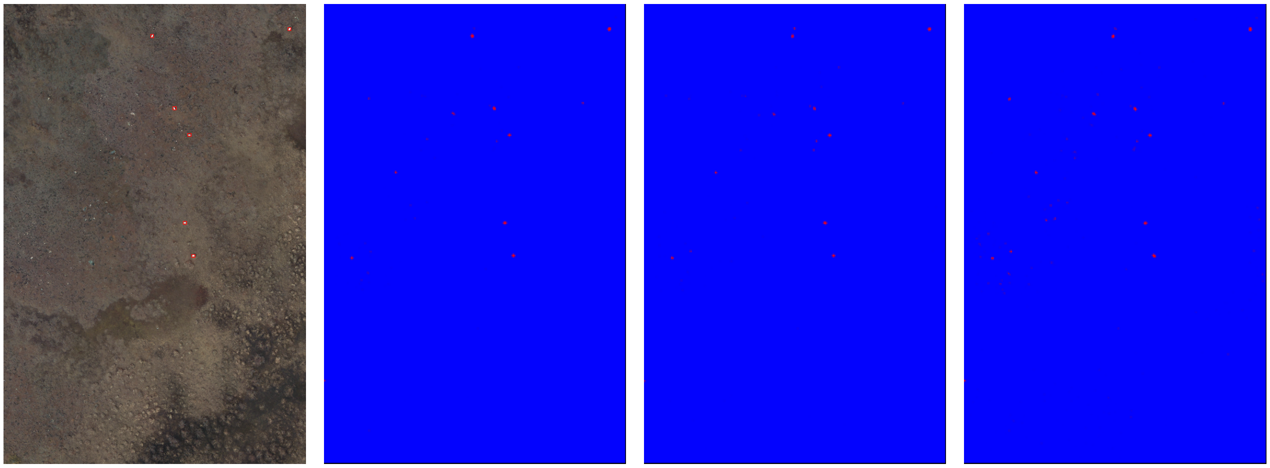
CNN	Matched	Missed	Total	Match Percent	BG Misclassified
Expert	375	26		93.52%	0.04%
Matched	360	41	401	89.78%	0.05%
Unmatched	379	22		94.51%	0.09%

identifying background is important. Considering the large amount of background and the fact there are no background "objects" like there are foreground objects, quantifying the accuracy on the background is more difficult than quantifying the accuracy on the foreground.

There are a few situations where humans are currently significantly better at identifying correctly than the CNNs developed in this paper. These cases, described below, accounted for many of the geese the CNNs failed to find and also for some of the misclassified background.

The process used to mosaic the images can create artifacting. Dramatic rainbow artifacting can occur on the right and bottom edge of the mosaics, as seen in Fig. 5. This dramatic coloration can confuse the CNN as there is the possibility of random coloration being similar to the objects being detected. However, this artifacting is obviously an error to human observers.

The best way to combat the artifacting issue would be to either: *(i)* manually crop the mosaics to ensure there are no artifacts in the dataset; or *(ii)* use the original images from the UAS flights instead of the mosaic images. The first solution is manageable on the small scale and is a reasonable preprocessing step before inclusion into the dataset, but may become cumbersome for larger projects and generalization. The second solution increases the number of images for identification by an order of magnitude, meaning the data will take longer to categorize by the citizen scientists and experts alike. The second solution also requires methods to handle



(a) Original Image (red boxes around expert identified white phase snow geese)

(b) Prediction by expert CNN

(c) Prediction by matched CNN

(d) Prediction by unmatched CNN

Fig. 6. Predictions on an image with rocks being misclassified as snow geese. The expert CNN misclassified the least amount of background with matched doing slightly worse and the unmatched doing worse still.

double counting in overlapping images.

Phantom objects are another potential error from the mosaicing process. Since the mosaic is generated by interleaving several images with an 80% overlap, sometimes the algorithm gets confused on the exact location of an object and includes a phantom object nearby. As the images are taken by a UAV over time, it is also possible that some geese are in some location one image, but by the time an overlapping image is taken they have moved. These phantom objects are more transparent than the original object, as seen in Fig. 7. Human observers can easily identify phantom objects; however, the CNNs have difficulty identifying phantoms due to the transparent nature of the phantom objects altering the RGB values significantly enough from a true object.

Another case where the CNNs was less likely to find a goose than a user is if part of the goose was hanging off the edge of the image. In this case, there might not be enough pixels comprising of the goose for the CNN to identify. Also most of the training sub-images have the geese roughly centered in the image.

2) Blob Counter: The prediction images had a blob detector run over them to estimate the number of each species in that image. The numbers generated by the blob detector were then compared to the number of geese identified by the experts. The results for each CNN over the whole dataset of test MSIs can be found in Table VI. It can be seen that the expert CNN did the best overall, but still over estimated the number of white phase snow geese by 100%. Although the unmatched dataset did the best at finding all the white phase geese identified by the experts as seen above, it did the worst in terms of overestimating the number of geese. Presumably, the reason for this is that the unmatched dataset is more likely to have rocks misidentified as white phase snow geese. Evidence of this hypothesis can be found in Figure 5 where it can be seen that the unmatched CNN identified the large rock in the image as white phase. The expert CNN, however, correctly identified it as background.

As noted above, some of the images contained artifacts from the mosaicing process. These accounted for a large of number of the blobs counted, especially when using the unmatched CNN. To see how badly artifacts affected the CNNs, the images containing artifacts were discarded from consideration in the blob counter. The results after removing artifacted images can be found in the "%Error Adjusted*" column of Table VI.

After removing MSIs that contained artifacts, 233 MSIs remained. For the expert dataset, the total error in the count was +331 blobs. 227 of these errors were found in only 23 images[6]. That is, almost 70% of the error was contained in approximately 10% of the images. For the matched dataset, approximately 77% of the errors were in 10% of the images and for the unmatched set 65% of the errors were in 10% of the images[7].

As seen by the results above, a small amount of background misclassification by the CNN in the training and testing phases can propagate into a large misclassification during the blob counter. Over the entire imagery, background makes up more than 99% of the total pixel area. It can be seen by comparing Tables V and VI that a mere increase of 0.05% of misclassified background pixels between the expert and unmatched dataset caused an extra 306% error in the blob counter. This is considerable, especially when it is taken into account that the expert CNN did only 0.16% better on the test IDX than the unmatched CNN. From the results of these networks,

[5] This is the error when MSIs containing artifacts were disregarded

[6] Errors were computed on a per image basis

[7] The particular images in the worst 10% varied from CNN to CNN with some overlap

TABLE VI
BLOB COUNTER RESULTS - WHITE PHASE ONLY

CNN	Calculated	Actual	Error	%Error	%Error Adjusted[5]
Expert	791		+397	+100%	+88%
Matched	1136	394	+742	+188%	+150%
Unmatched	1993		+1599	+406%	+250%

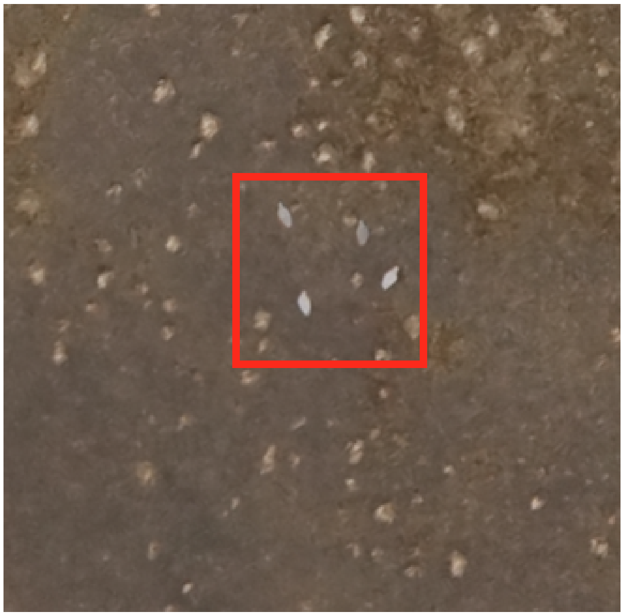


Fig. 7. Example of phantom objects created by the mosaicing process

it would seem that it is desirable to be a little better at classifying background at the cost of being not as good at identifying geese. This is, once again, because of the nature of the data, massive amounts of background with tiny amounts of foreground.

VI. CONSLUSION

While the results are encouraging, especially when compared against similar automated object detection projects using RGB images from UAS flights, there are still several issues that need addressing to reliably run a CNN over the UAS imagery to accurately detect lesser snow geese — and potentially other ecological data. The extremely small size of the objects compared to the background, the minuscule ratio of foreground to background in the images, and the prevalence of background with similar characteristics to the objects being detected are all major concerns.

Although there is still some work to be done to make the CNNs more accurate, there were encouraging results showing the efficacy of using citizen science observations as input for the CNNs. As shown in the blob counter results, Sec. V-B2, individual citizen scientists produce training data that is a poor candidate for CNN training when compared to experts. However, when matching individual citizen scientists and taking the intersecting observations of the same objects, the quality of the training data is dramatically improved. The matched observations are still not as good as the expert data, but with further matching to more than two users and implementation of some techniques discussed below, citizen scientists can be further shown to be useful for creating good initial training data.

What do these results mean for automation of detection and counting of lesser snow geese? Even if the blob counter isn't perfect in accuracy, the CNNs can still save a significant amount of expert and citizen scientist time. Instead of showing every image to the users, only images with areas identified by the CNN as non-background could be shown to the users. In addition to reducing the amount of imagery shown it could potentially help reduce errors by the users. By mostly showing "interesting" images containing geese and reducing the number of uninteresting, monotonous images shown, user fatigue could potentially be reduced. Testing whether this conjecture holds true is outside of the scope of this paper, but its potential benefit should not be overlooked.

VII. FUTURE WORK

A. Feedback Loop

There are some similar objects that are misclassified throughout the imagery. One proposal to decrease the amount of false positives is to ensure that the commonly misclassified objects are included in the background dataset, thus increasing the probability of being correctly identified. To automate this process, a feedback loop will be implemented that takes the misclassified areas from the blob counter as shown in Sec. V-B2 and retrains the CNN on the "harder" data to identify. It is hopeful that this process can be fully automated and reduce the misclassification errors of the CNNs dramatically. An initial exploration into this has been started and is giving promising results.

B. Turn-Key Application

An obstacle for any field-scientist wanting to implement a similar system will be the requirement of a technical team to develop and maintain the web portal and infrastructure. As a continuation of the citizen science portion of this project, an attempt to generalize the architecture to develop a turn-key system will be made. The goal is that this system will be configurable and usable by a variety of projects requiring object detection in imagery. The generalized system would still require a technical team to get running, but should increase the viability of using these techniques and systems on other similar projects.

ACKNOWLEDGMENTS

We appreciate the support and dedication of the Wildlife@Home citizen scientist, and Jennifer Booth specifically, who have spent significant amounts of time reviewing images and classifying objects within them.

Funding was provided by North Dakota EPSCoR, the Hudson Bay Project, Central and Mississippi Flyways, and the UND College of Arts and Sciences. UAS data collection supported by the Hudson Bay Project.

Special thanks to M. P. Corcoran for assistance in flight operations.

REFERENCES

[1] M. Mattingly, A. Barnas, S. Ellis-Felege, R. Newman, D. Iles, and T. Desell, "Developing a citizen science web portal for manual and automated ecological image detection," in *e-Science (e-Science), 2016 IEEE 12th International Conference on*. IEEE, 2016, pp. 223–232.

[2] C. J. Lintott, K. Schawinski, A. Slosar, K. Land, S. Bamford, D. Thomas, M. J. Raddick, R. C. Nichol, A. Szalay, D. Andreescu, P. Murray, and J. Vandenberg, "Galaxy zoo: morphologies derived from visual inspection of galaxies from the sloan digital sky survey," *Monthly Notices of the Royal Astronomical Society*, vol. 389, no. 3, pp. 1179–1189, 2008.

[3] C. Lintott, K. Schawinski, S. Bamford, A. Slosar, K. Land, D. Thomas, E. Edmondson, K. Masters, R. C. Nichol, M. J. Raddick, A. Szalay, D. Andreescu, P. Murray, and J. Vandenberg, "Galaxy zoo 1: data release of morphological classifications for nearly 900,000 galaxies," *Monthly Notices of the Royal Astronomical Society*, vol. 410, no. 1, pp. 166–178, 2011.

[4] J. e. a. Adelman-McCarthy, "The 6th Sloan Digital Sky Survey Data Release, http://www.sdss.org/dr6/," July 2007, apJS, in press, arXiv/0707.3413.

[5] D. A. Fischer, M. E. Schwamb, K. Schawinski, C. Lintott, J. Brewer, M. Giguere, S. Lynn, M. Parrish, T. Sartori, R. Simpson, A. Smith, J. Spronck, N. Batalha, J. Rowe, J. Jenkins, S. Bryson, A. Prsa, P. Tenenbaum, J. Crepp, T. Morton, A. Howard, M. Beleu, Z. Kaplan, N. vanNispen, C. Sharzer, J. DeFouw, A. Hajduk, J. P. Neal, A. Nemec, N. Schuepbach, and V. Zimmermann, "Planet hunters: the first two planet candidates identified by the public using the kepler public archive data," *Monthly Notices of the Royal Astronomical Society*, vol. 419, no. 4, pp. 2900–2911, 2012.

[6] Lion Research Center, University of Minnesota, [Accessed Online, 2012] http://www.snapshotserengeti.org/.

[7] R. Simpson, K. R. Page, and D. De Roure, "Zooniverse: observing the world's largest citizen science platform," in *Proceedings of the 23rd international conference on world wide web*. ACM, 2014, pp. 1049–1054.

[8] J. Cox, E. Y. Oh, B. Simmons, C. Lintott, K. Masters, A. Greenhill, G. Graham, and K. Holmes, "Defining and measuring success in online citizen science: A case study of zooniverse projects," *Computing in Science & Engineering*, vol. 17, no. 4, pp. 28–41, 2015.

[9] R. Bonney, C. B. Cooper, J. Dickinson, S. Kelling, T. Phillips, K. V. Rosenberg, and J. Shirk, "A developing tool for expanding science knowledge and scientific literacy," *BioScience*, vol. 59, pp. 977–984, 2009.

[10] M. A. Voss and C. B. Cooper, "Using a free online citizen-science project to teach observation and quantification of animal behavior," *American Biology Teacher*, vol. 72, pp. 437–443, 2012.

[11] ——, "Using a free online citizen-science project to teach observation & quantification of animal behavior," *The american biology Teacher*, vol. 72, no. 7, pp. 437–443, 2010.

[12] C. Wood, B. Sullivan, M. Iliff, D. Fink, and S. Kelling, "ebird: engaging birders in science and conservation," *PLoS biology*, vol. 9, no. 12, p. e1001220, 2011.

[13] L.-P. Chrétien, J. Théau, and P. Ménard, "Visible and thermal infrared remote sensing for the detection of white-tailed deer using an unmanned aerial system," *Wildlife Society Bulletin*, vol. 40, no. 1, pp. 181–191, 2016.

[14] A. Abd-Elrahman, L. Pearlstine, and F. Percival, "Development of pattern recognition algorithm for automatic bird detection from unmanned aerial vehicle imagery," *Surveying and Land Information Science*, vol. 65, no. 1, p. 37, 2005.

[15] D. B. Sasse, "Job-related mortality of wildlife workers in the united states, 1937-2000," *Wildlife society bulletin*, pp. 1015–1020, 2003.

[16] M. Israel, "A uav-based roe deer fawn detection system," in *Proceedings of the International Conference on Unmanned Aerial Vehicle in Geomatics (UAV-g), H. Eisenbeiss, M. Kunz, and H. Ingensand, Eds*, vol. 38, 2011, pp. 1–5.

[17] B. E. Wilkinson, "The design of georeferencing techniques for unmanned autonomous aerial vehicle video for use with wildlife inventory surveys: A case study of the national bison range, montana," Ph.D. dissertation, University of Florida, 2007.

[18] R. P. Breckenridge, M. Dakins, S. Bunting, J. L. Harbour, and S. White, "Comparison of unmanned aerial vehicle platforms for assessing vegetation cover in sagebrush steppe ecosystems," *Rangeland Ecology & Management*, vol. 64, no. 5, pp. 521–532, 2011.

[19] W. Koski, P. Abgrall, and S. Yazvenko, "A review and inventory of unmanned aerial systems for detection and monitoring of key biological resources and physical parameters affecting marine life during offshore exploration and production activities," *IWC paper SC/61 E*, vol. 9, 2009.

[20] P. Soriano, F. Caballero, A. Ollero, and C. A. de Tecnologıas Aeroespaciales, "Rf-based particle filter localization for wildlife tracking by using an uav," in *On: 40th International Symposium of Robotics, Barcelona, España*, 2009.

[21] M. Leonardo, A. M. Jensen, C. Coopmans, M. McKee, and Y. Chen, "A miniature wildlife tracking uav payload system using acoustic biotelemetry," in *ASME 2013 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*. American Society of Mechanical Engineers, 2013, pp. V004T08A056–V004T08A056.

[22] R. P. Breckenridge, R. Lee, A. Piscitella, and C. Eckersell, "Using unmanned aerial vehicles to assess vegetative cover in sagebrush steppe ecosystems," Idaho National Laboratory (INL), Tech. Rep., 2005.

[23] J. Linchant, C. Vermeulen, J. Lisein, P. Lejeune, and P. Bouché, "Using drones to count the elephants: a new approach of wildlife inventories," 2013.

[24] K. A. Steen, A. Villa-Henriksen, O. R. Therkildsen, H. Karstoft, O. Green *et al.*, "Automatic detection of animals using thermal imaging," in *International Conference on Agricultural Engineering*, 2012.

[25] C. Cariappa, W. Ballard, S. Breck, A. J. Piaggio, and M. Neubaum, "Estimating population size of mexican wolves noninvasively (arizona)," 2008.

[26] C. Vermeulen, P. Lejeune, J. Lisein, P. Sawadogo, and P. Bouché, "Unmanned aerial survey of elephants," *PloS one*, vol. 8, no. 2, pp. 1–7, 2013.

[27] G. Caughley, "Bias in aerial survey," *The Journal of Wildlife Management*, pp. 921–933, 1974.

[28] ——, "Sampling in aerial survey," *The Journal of Wildlife Management*, pp. 605–615, 1977.

[29] Y. LeCun and C. Cortes, "Mnist handwritten digit database," *AT&T Labs [Online]. Available: http://yann. lecun. com/exdb/mnist*, 2010.

[30] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision (IJCV)*, vol. 115, no. 3, pp. 211–252, 2015.

[31] C. Bowley, A. Andes, S. Ellis-Felege, and T. Desell, "Detecting wildlife in uncontrolled outdoor video using convolutional neural networks," in *e-Science (e-Science), 2016 IEEE 12th International Conference on*. IEEE, 2016, pp. 251–259.

[32] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International Conference on Machine Learning*, 2015, pp. 448–456.

[33] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. ICML*, vol. 30, 2013, p. 1.