Utilizing Neural Networks to Resolve Individual Bats and Improve Automated Counts

Ian Bentley
Department of Engineering Physics
Florida Polytechnic University
Lakeland, FL, USA
ibentley@floridapoly.edu

Marwan Gebran
Department of Chemistry and Physics
Saint Mary's College
Notre Dame, IN, USA
mgebrane@saintmarys.edu

Sam Vorderer
Department of Chemistry and Physics
Saint Mary's College
Notre Dame, IN, USA
svorderer01@saintmarys.edu

Joel Ralston
Department of Biology
Saint Mary's College
Notre Dame, IN, USA
jralston@saintmarys.edu

Laura Kloepper
Department of Biological Sciences
University of New Hampshire
Durham, NH, USA
laura.kloepper@unh.edu

2023 IEEE World AI IoT Congress (AIIoT) | 979-8-3503-3761-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/AIIoT58121.2023.10174554

***Abstract*—Accurate population counts are essential for understanding the status of species and for researchers studying various phenomena including monitoring the relationship between environmental stresses and the spread of disease within populations. Both small roosts and large colonies of bats provide challenges when attempting to determine an accurate population count. Recently, there have been a number of new video analysis software applications, that are available on the internet, which can be used to provide population counts. When software-based counts are compared with manual counts, the software provides counts that are substantially less labor intensive, determined substantially more quickly, and have the potential to be more accurate. This short paper discusses the use of neural networks to determine the number of bats that there are in a region when multiple bats may overlap. The work discussed in this manuscript demonstrates that the counts of multiple overlapping bats can be improved using trained neural networks. This is a critical improvement for providing accurate counts in high density videos. This manuscript contains the biological motivations, and a brief overview of how artificial intelligence is being implemented. The results discussed compare the accuracy values of neural networks for a few case studies including cross-comparisons of data trained on different video types and for different animals which can have accuracy values above 90 % for comparable video types. Finally, the generation and use of synthetic images, to increase the amount of data in a training set, is also discussed, which resulted in a trained neural network that produced an accuracy value of 80% on 12 unbiased categories.**

***Index Terms*—artificial intelligence, bioinformatics, computer vision, machine learning, neural networks.**

I. Introduction

Bats are found worldwide, are vital to healthy ecosystems, and provide crucial benefits to human systems globally through seed dispersal, crop pollination, and reduction of agricultural pest and insect disease vectors [1]- [3]. Yet the current population status of many bat species is unknown because they are often densely populated, mobile, and roost in difficult to access locations, making them difficult to count. Monitoring the size of bat populations is important for assessing the need for management and conservation intervention, calculating changes in life history parameters, and monitoring species response to threats such as habitat destruction, climate change, and emerging diseases such as white-nose syndrome, see e.g., [4].

Recent advances in video technology, including near- and far-infrared, have increased our capacity to monitor bats emerging from roosts, and many agencies use these cameras to record bats. Until recently, many roost counts were conducted by visual estimates assuming constant flow rate during an emergence, or by manually counting individuals from video recordings. However, these methods are very time intensive, and can be inaccurate and prone to bias, see e.g., [5].

One approach for determining the count of bats in a video is to use automated video analysis software which can be downloaded from the internet (see e.g., [6] and [7]). A description of how both of the above software packages can be used for this purpose is described in Refs. [8] and [9]. In these software packages, moving objects are identified and counts are determined for bats that enter or exit a boundary. The videos analyzed in Ref. [9] demonstrate that the automated counts were on average 83% of the value of the rigorous manual counts across multiple video segments. In every case described in that analysis, the automated count was below the manual count. This is in large part because the software counted multiple overlapping bats as a single bat. The neural network techniques explored in this manuscript will improve the resulting counts by resolving the actual number of objects leaving a boundary.

Instead of looking for specific features to aid the count, we will use neural networks. Neural networks are a mathematical tool that can be used to categorize objects. This is used in both self driving and driving assisted vehicles to identify other vehicles, road markings, road signs, and obstacles, see

This work was supported by the National Science Foundation (Award Number 1916850).

979-8-3503-3761-7/23/$31.00 ©2023 IEEE

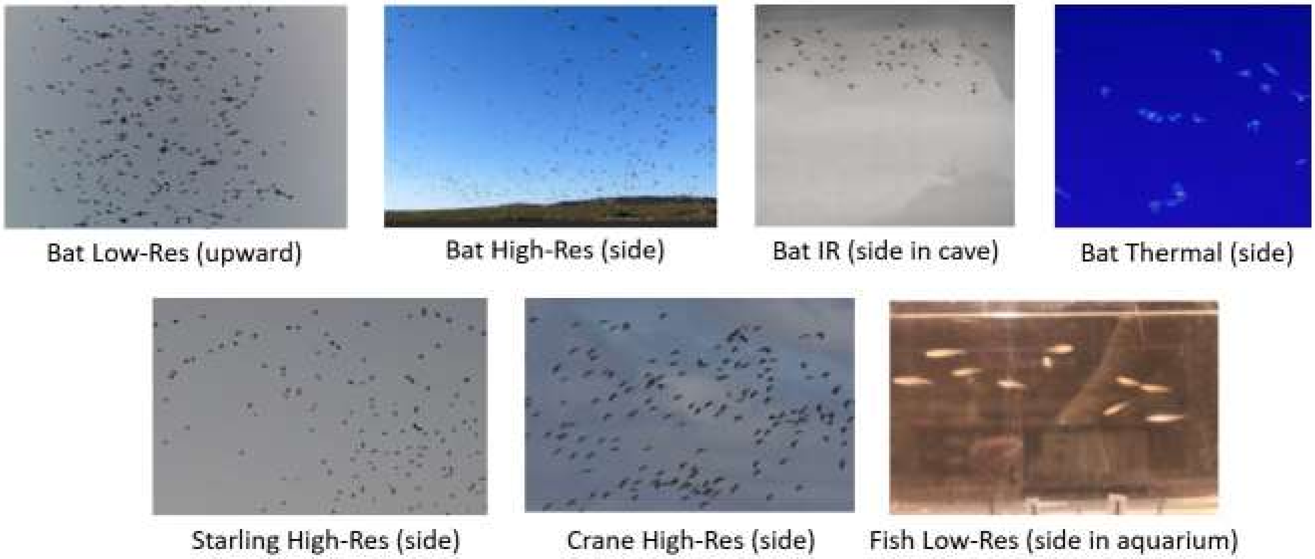

Fig. 1. Full frame snapshots for the seven videos used to train neural networks. The top row contains videos of bats from different sources. The bottom includes two bird videos and the last is of a fish video.

e.g., [10]. It is also used in handwriting recognition in an automated teller machine to read the amount of money that should be transferred, see e.g., [11]. In the work described in this manuscript, we will train, test, and ultimately implement a neural network for use in software to provide more accurate counts of overlapping bats in videos.

There are multiple video types currently used in population studies of bats including, low resolution GoPro videos, higher resolution videos, infrared videos, and thermal videos. We will test each of these and see if one neural network can be used for all video types or if individual networks are required. Further, we will also test if other species (specifically birds and fish), can also be accurately counted using a universal network trained from the multiple input data sets.

II. Data preparation for network training

Seven test videos are discussed in this manuscript which were used to train the counts. These videos include four of bats, two of birds, and one with fish. Figure 1 includes a single frame from these seven videos. These different videos are representative of different field studies beyond those with the goal of determining bat populations. This set of videos will be used to test the cross-applicability of the trained neural networks.

A. Determining regions of interest and the counts within them

All seven videos discussed in this manuscript were taken from a stationary perspective. That allows for the background to be determined based the median of pixel value for each stack of video frames [12]. The background from each video was then subtracted from each individual frame leaving foreground objects. A bounding box was placed around all foreground objects that were detected. This bounding box defined the separate regions of interest that were then categorized and used to train neural networks.

In the analysis described in this manuscript, 12 categories have been used corresponding to number of animals (bats, birds, or fish) found in a region of interest. These count categories range from zero to ten, with one additional category that include all the boxes containing more than 10 animals. The frequencies at which this highest count (10+) occurred was less than 0.25% of that count for each of the seven videos.

Our target was to categorize 10,000 counts for each video to be used in training and validation. To facilitate this task, a graphical user interface was developed which allowed us to label each region of interest with the corresponding count as well as to add a flag if this number was uncertain. After labels were assigned the next region of interest was cycled to.

After eliminating all uncertain counts, a total count of 29,732 bats, 18,057 birds, and 9,956 fishes were categorized across seven videos. Table I contains the raw counts for each category across the seven videos as well as the percentage that each count has. The total count of the thermal video was lower than our target of 10,000 because it had fewer frames resulting in too few regions of interest to categorize.

A low resolution GoPro video of bats against a grey sky at dusk will be the primary focus of the discussion that follows. This video will be used to demonstrate the need, creation, and quality of synthetic images. In Fig. 2, a sample of the region of interests for the each of the 12 categories is shown.

B. Distributions of counts in regions of interest

The counts depicted in Tab. I are heavily biased toward the first three categories (0, 1, and 2 counts). These categories account for more than 96% of the overall counts made in the seven original videos. The presence of the high number of zero counts in some videos was caused by the background detection

TABLE I

OCCURRENCE AND PERCENT FOR EVERY COUNT CATEGORY (0-10+) FOR THE SEVEN INDIVIDUAL VIDEOS AND ALL SEVEN VIDEOS COMBINED.

Video	0	1	2	3	4	5	6	7	8	9	10	10+	Total
Low-Res Bat	645 *6.94%*	6376 *68.6%*	1503 *16.2%*	368 *3.96%*	201 *2.16%*	87 *0.94%*	50 *0.54%*	31 *0.33%*	5 *0.05%*	7 *0.08%*	12 *0.13%*	6 *0.06%*	9291
High-Res Bat	1 *0.01%*	7711 *92.4%*	509 *6.10%*	90 *1.08%*	27 *0.32%*	6 *0.07%*	2 *0.02%*	0 *0.00%*	0 *0.00%*	0 *0.00%*	0 *0.00%*	0 *0.00%*	8346
IR Bat	236 *2.43%*	7395 *76.0%*	1306 *13.4%*	390 *4.01%*	168 *1.73%*	101 *1.04%*	47 *0.48%*	26 *0.27%*	15 *0.15%*	17 *0.17%*	4 *0.04%*	23 *0.24%*	9728
Thermal Bat	2 *0.08%*	2234 *94.4%*	118 *4.99%*	11 *0.46%*	1 *0.04%*	1 *0.04%*	0 *0.00%*	0 *0.00%*	0 *0.00%*	0 *0.00%*	0 *0.00%*	0 *0.00%*	2367
High-Res Starling	1 *0.01%*	8208 *91.8%*	640 *7.16%*	60 *0.67%*	16 *0.18%*	2 *0.02%*	0 *0.00%*	0 *0.00%*	0 *0.00%*	1 *0.01%*	0 *0.00%*	10 *0.11%*	8938
High-Res Crane	0 *0.00%*	7906 *86.7%*	921 *10.1%*	213 *2.34%*	40 *0.44%*	33 *0.36%*	5 *0.05%*	1 *0.01%*	0 *0.00%*	0 *0.00%*	0 *0.00%*	0 *0.00%*	9119
Low-Res Fish	7711 *77.5%*	2086 *21.0%*	95 *0.95%*	5 *0.05%*	2 *0.02%*	0 *0.00%*	0 *0.00%*	1 *0.01%*	4 *0.04%*	28 *0.28%*	23 *0.23%*	1 *0.01%*	9956
All 7 Combined	8596 *14.9%*	41916 *72.6%*	5092 *8.82%*	1137 *1.97%*	455 *0.79%*	230 *0.40%*	104 *0.18%*	59 *0.10%*	24 *0.04%*	53 *0.09%*	39 *0.07%*	40 *0.07%*	57745

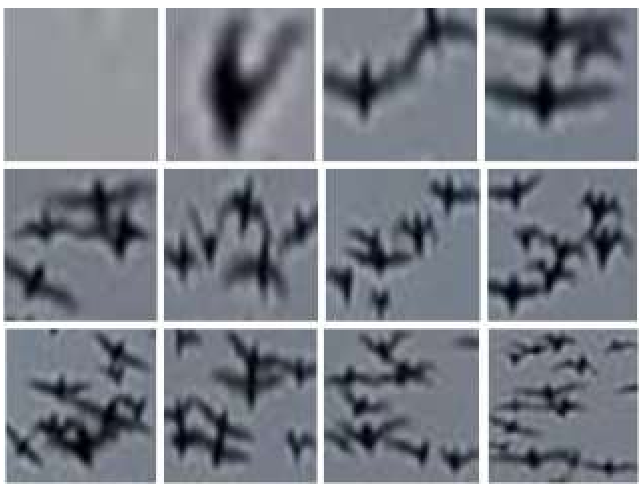

Fig. 2. Sample snapshots from small regions in a GoPro video demonstrating the 12 categories that are used in training. Each category is defined by the number of bats seen. The number of bats, from left to right, are: (top row) 0, 1, 2, 3, (middle row) 4, 5, 6, 7, and (bottom row) 8, 9, 10, and 10+.

being sensitive to fluctuations in lighting, the movement of a background object (e.g. a cloud, or a tree), or a cave edge. In the case of the fish video, it was glare and reflection on the glass of the aquarium that created multiple regions with no detected object inside. In the fish video, the zero count was actually the most frequently occurring category accounting for more than 77% of the counts in the video.

In each of the other videos the one count category occured most commonly. In all videos, the density of the bats, birds, or fish, was low enough that more often than not, single objects were found in the regions of interest as opposed to multiple overlapping objects. Table I indicates that in many cases, for example the Thermal Bat video in categories with counts greater than three, there is a dearth of data.

The count of the number of bats in the GoPro video was performed by three people. These counts were done for the same regions of interest, on the same video. Overall there was a greater than 99% agreement among these separate count catalogs.

The issue of training on an imperfect data set that is not uncommon in classification problems [13]. The effect of having noisy or mislabelled data is explained in detail in [14] and is negligible. This is especially true when considering that we will incorporating techniques such as data augmentation and regularization, in our final neural network that will be implemented in tracking software. These will help improve the network's robustness to mislabelled data [15], [16].

III. DEEP LEARNING ANALYSIS

Counting bats, birds, or fishes is a multi-class classification problem. We have chosen to apply deep neural network techniques in order to classify the number of bats in each image between 0 and 11. The network takes as input an RGB image of dimension (40,40,3) and outputs a probability distribution among the 11 classes. This image dimension was chosen because it didn't involve substantial up-scaling of low count regions or down-scaling of high count regions. All regions of interest were resized to match this common dimension. All images pass through a normalisation procedure before entering the network. The purpose of this step is to transform the values of the pixels from (0,255) to (0,1). Two types of normalization were tested and found to give similar results in terms of network accuracy. The first is done by dividing all pixels by 255 and the second is by performing a pixel standardization which is done by removing, from each image, the mean of the dataset and dividing it by the standard deviation of the dataset.

Many parameters were optimized before the construction of the network. Initializers, optimizers, learning rates, dropout fraction, pooling layers, activation functions, loss functions, epochs, and batches were each constrained in an iterative manner and the best architecture was selected according to its simplicity (size and calculation time) and to the accuracy of the results. Fully dense NNs, Convolutional Neural Networks

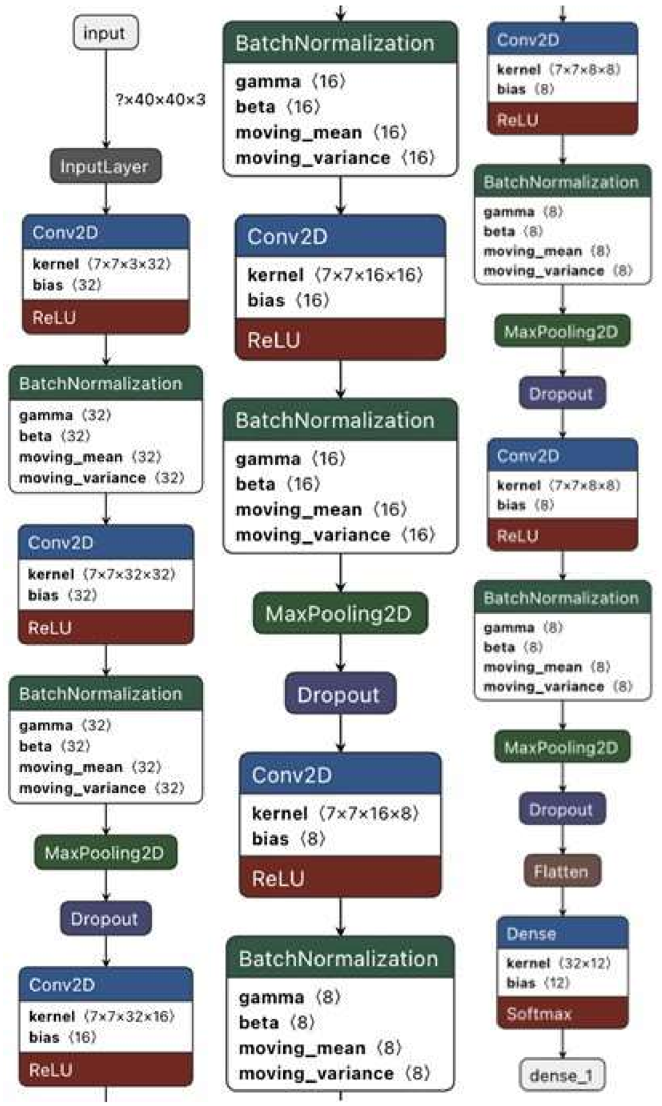

Fig. 3. Architecture of the neural network used to categorize 12 counts for identified regions of interest. The architecture continues from the bottom of the first column to the top of the second, and similarly, the second column is continued on the top of the third column.

(CNN), and a combination of both were tested. The number of hidden layers, the number of neurons in each layer, and the size of the filters are derived for each architecture. We also used the `KerasTuner` package for this optimization. `KerasTuner` is a hyperparameter optimization framework that aims to alleviate the challenges of searching for optimal hyperparameter values. After testing various networks, the optimized neural network is displayed in Fig. 3 was decided on. For each video we used 80% of the data set for training and the remaining 20% was set aside for an initial validation.

IV. Results

The discussion below has been broken into three subsections. Section IV-A will discuss how well each network was trained and how well it worked when applied to a different video. Section IV-B describes the use of a network trained from the combined data set from all seven videos. Section IV-C discusses the creation and use of synthetic images that were used to supplement the analysis of the low resolution GoPro video.

A. Evaluating networks trained on different data sets

The diagonal (top-left to bottom-right) entries in Table II show that training a network on a specific type of images results in a high accuracy when tested on the same type of images. In those cases, the classification accuracy ranged from 95-99%.

It should be stated that often in these cases, guessing a count of one object or zero objects would more often than not be correct. Nonetheless, the determined accuracy values are promising, especially when considering that 12 categories were used. Further, in looking at the confusion matrices for these self-compared networks often when a count was off, it was often off only slightly (over or under estimating by only one or two). For our purpose of training a network to improve counts of overlapping bats this will be very beneficial. This is especially true when considering the large statistics involved in a population count, were a handful of over and under estimates can largely cancel out.

Table II can be used for cross-comparisons as well. In this table the rows represent the training data set used to create the corresponding network and the columns contain what regions of interest were counted using that network. So for example the top row indicates that the network trained on the low-resolution GoPro video was somewhat accurate when applied to the Bat IR and Starling videos (75% and 69%, accurate respectively). But it had a lower accuracy for the other four cross-comparisons. This can be reasoned out by the fact that aesthetically the Bat IR and Staring videos, also had a grey background most similar to the trained video, and the size and shape of the bats (or birds) were also comparable to the objects in the trained video. The Bat Thermal video and Crane videos resulted the least accurate use of this network, potentially because of the different backgrounds in the prior and the different object shapes in the latter.

Most networks worked decently well (>70% accuracy) when applied to one or two other videos, and poorly for the remaining. The network trained on the Starling video data was arguably the most universal, working well for all data sets except for the Fish video data. In all fairness, this is likely because in many quantifiable measures the Starling video data is similar to that of the other videos it compared well with. Not surprisingly, the Fish video data, being most dissimilar from the other videos in object shape, object color, and background resulted in the least cross applicable network resulting in an accuracy ranging from 7%-36%.

B. Training and testing one network applied to all videos

When implementing the neural network to provide more accurate counts of multiple overlapping objects a natural question to ask is whether or not separate networks will need to be trained for each video type, each species, and so on.

TABLE II
VALIDATION AND CROSS-COMPARISONS OF NEURAL NETWORKS FROM DIFFERENT VIDEOS.

	Low-Res Bat	High-Res Bat	IR Bat	Thermal Bat	High-Res Starling	High-Res Crane	Low-Res Fish	All 7 Combined	Low-Res Synthetic
Low-Res Bat	98%	17%	75%	3%	69%	9%	30%	48%	24%
High-Res Bat	73%	97%	76%	77%	96%	83%	19%	73%	10%
IR Bat	11%	10%	95%	0%	5%	4%	71%	33%	18%
Thermal Bat	69%	88%	76%	98%	80%	82%	21%	70%	8%
High-Res Starling	69%	92%	78%	94%	99%	89%	0%	71%	8%
High-Res Crane	69%	71%	76%	2%	94%	99%	21%	68%	8%
Low-Res Fish	18%	21%	7%	23%	22%	36%	96%	34%	12%
All 7 Combined	92%	95%	88%	95%	98%	95%	94%	94%	24%
Low-Res Synthetic Bat	80%	22%	35%	0%	34%	9%	77%	42%	63%

To investigate this we attempted to make a more universal network comprised of the pooled data from all 57,745 counted objects across the seven videos. An eighth (All 7 Combined) network was trained based on this data set.

Still keeping in mind the heavy bias toward 0-3 objects, the results are on the second to last row of Table II. The overall accuracy is 94%. This network was also tested against each data set individually. The accuracy values range from 88% for the IR Bat data up to 98% for the Starling data. For each comparison, the overall network sacrifices some accuracy but it has the potential to be universally applied.

C. Generating synthetic images

Across all seven videos most of our regions of interest contain 0-3 detected objects. This has created a bias in the network which will prefer lower counts. This is also a problem if there isn't sufficient data (or any data at all) to properly train a count category for a specific network. Additionally, even when there is some data to train on there may be less than it initially appears.

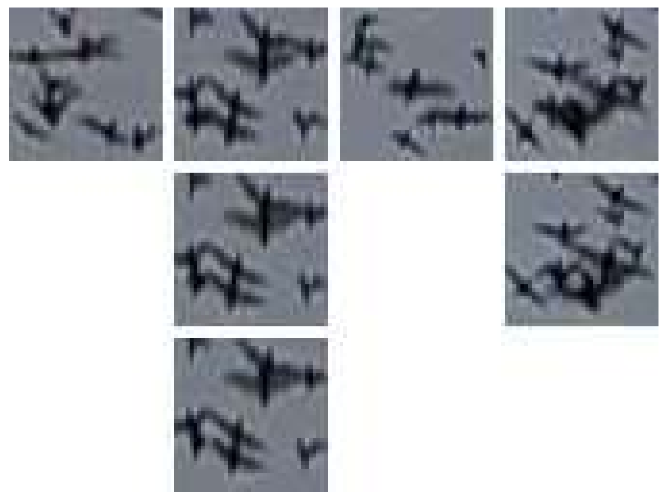

Fig. 4. The seven regions of interest that have been categorized as having nine bats inside. These have been arranged into columns to highlight two occurrences of consecutive frames with minimal variations within.

Data augmentation (specifically, applying skew and/or rotation transforms) is usually used to generate additional images and increase the volume of the data sets and to remove any existing bias among categories [17]. This process acts as a regularizing technique and helps in avoiding over-fitting. Take, for example, the category of nine bats counted in the low resolution GoPro video, according to Table I there are seven different identified regions of interest with nine bats. Figure 4 shows that these seven regions can be reduced to four because some are based on the same arrangement seen in a subsequent frame. Overall, even with 10,000 counts categorized there is a lack of sufficient data for properly training a neural network in high count categories.

We have developed a simple algorithm to combat this lack of data and have used it to create synthetic images. The regions of interest from the low-resolution GoPro video were used to prototype this algorithm.

The first step is to use regions of interest with no object detected to define a background on which an object (a bat in this case) can be placed. The background determination was done in two ways. The first method for determining backgrounds is to use the bounding box locations from specific frames placed on the median determined background where no objects exist. The second method for determining a background consists of using the object detection with a very low tolerance on each frame so that background and foreground objects could both be found. The regions of interest that resulted were compared with the previously determined regions. All cases with no overlap (meaning no object was found inside) were saved and as new background regions. The first background method has the benefit of being regions where genuinely bats were seen, and the second has the benefit of potentially having additional background texture.

On each randomly generated background, it is possible to place the desired number of bats (from zero to more than ten). When placing a bat we first need to have a trusted set of regions of interest where we confidently know the count. To simplify matters we choose to use only images with a single bat. The 6,376 regions with a single bat were filtered so that there was no partial bat, boundary of the video, or anything else in the image that might also be copied. Those restrictions reduced our number of usable single bat regions by about a

Fig. 5. Regions of interest containing single bats. Original regions of interest with a single bat from the low resolution GoPro video are in the top row. Slightly augmented single bat regions of interest with a minor skew and rotation are in the middle row. Synthetic images consisting of a single bat from one region of interest placed on a background from another region on the video are in the bottom row.

Fig. 6. Twelve synthetic frames that have nine bats placed in random locations on a random background. These are intended to simulate and create additional data similar to what is shown in Fig. 4.

factor of four, down to 1,460.

The backgrounds that were previously discussed were used at random to generate a set of regions with zero bats. Then using the single bats, a broad portion of background was used and then one or more bat was placed on that background at random locations. A Gaussian filter is then used to blur the edges of the foreground object with the background to create the synthetic composite image.

Figures 5 and 6 have been included to show some of the synthetic images that have been generated. In Fig. 5 the top row displays the original frames with a single bat region of interest, the middle row displays a single bat region which has been slightly transformed (minor rotations or skew adjustments) and the final row displays single bats which have been placed on a single background.

In Fig. 6 multiple bats have been stacked on the image at different locations. When compared to Fig. 4, the synthetic regions have the benefit of being more randomly placed, and the additional benefit of having exactly the desired number of objects in it. In Figure 6 the spacing of the bats overall appears to be larger than in Figure 4.

A network was also trained for this completely synthetic data set created using GoPro video for background and the single bats. An unbiased data set comprised of approximately 6,000 synthetic images in each count category. The bottom row of Tab. II contains the results comparing this synthetic set with the others. The validation accuracy of this network was 63%. This number may sound poor when compared to prior results, but it is comprised of 12 count categories each with an equal weight. Further, we believe that, improvements to how the unbiased data set has been made, an additional augmentation step, and modifications of the neural network architecture will allow for a higher accuracy.

The most important result of using this synthetic network comes from applying the synthetically trained network with the original low resolution video. In that case, the synthetic trained network gave an 80% accuracy. The confusion matrix with the application of the synthetic data trained network applied to the original low resolution GoPro data set has been included as Figure 7.

This figure shows that the majority of the incorrect counts are off by a few values from the actual. It should however be noted that often an underestimate of the count was predicted by this network. This can be addressed by properly spacing the bats so that the network can better resolve multiple bats. Ultimately, the maximum range and minimum range of placement needs to be further tested. Nonetheless, we believe that the routine described here can be beneficial for training similar data sets where data is biased and otherwise sparse.

V. Ongoing and Future Work

In future work we will test maximum and minimum spacing for subsequent bats in the hopes of allowing all bats to be at least partially seen while still having them placed close enough together that they will be identified as a single object by an untrained eye.

Additionally, because high counts (more than five) occur so infrequently we will also train future networks with fewer categories and we will use additional data augmentation techniques to further increase our training pool. These can simply include using the skew and rotation augmentation on synthetically generated images.

Another question of interest is whether we can use the synthetically generated images where we know each bat's center to provide a better tracking algorithm. We plan to test this in the near future on synthetic images with a few (five or fewer) bats.

We will need to run this augmentation routine on each of the different video types. The low resolution GoPro video was specifically chosen because of the ease in determining foreground objects and because of its simple background. The IR video for example with the cave edges in the background

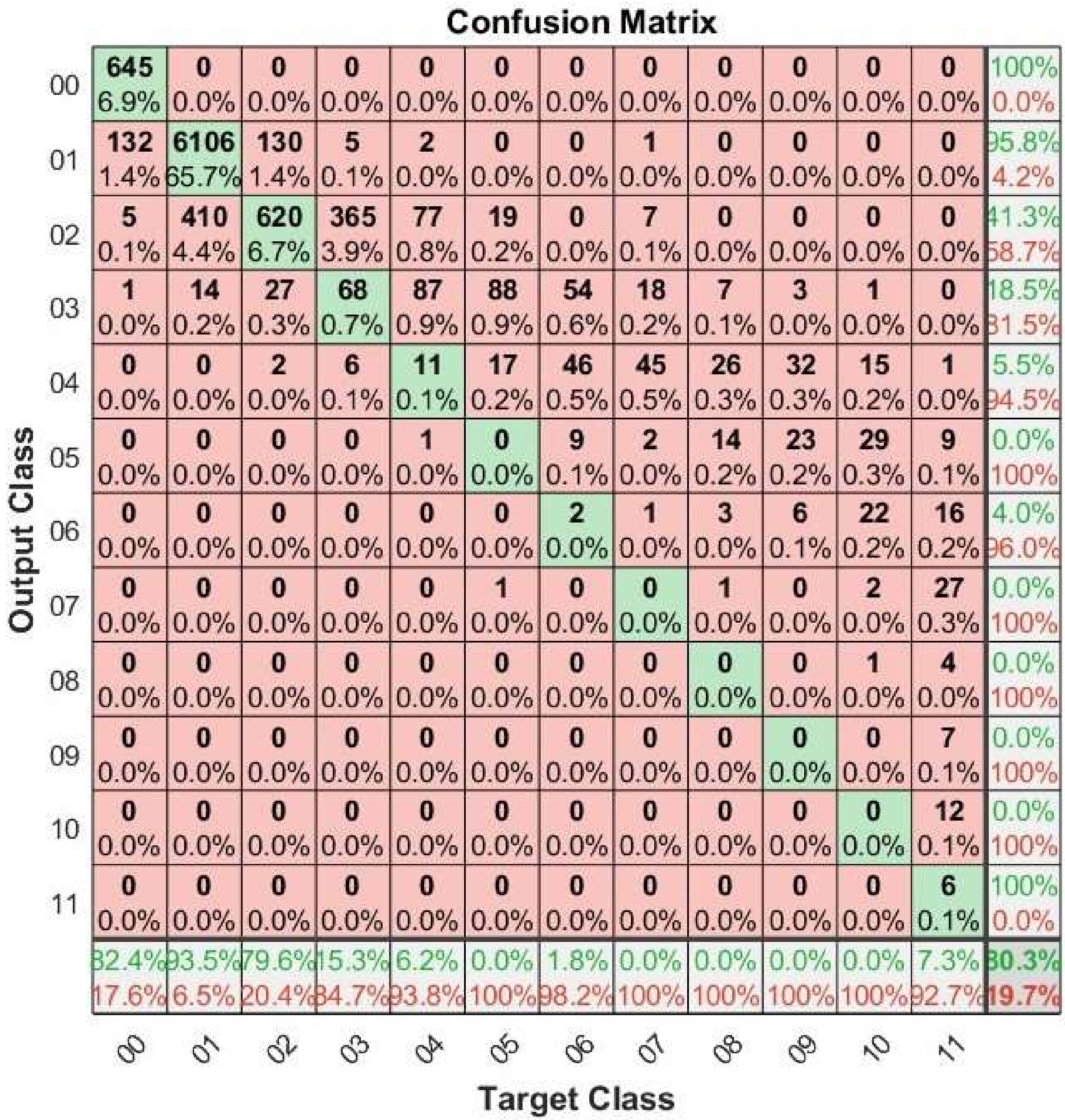

Fig. 7. Confusion matrix with the occurrences of predicted counts (rows) against actual counts (columns) based on a purely synthetic data trained network used to classify the original GoPro images. The synthetic data was generated from stacking low resolution GoPro images of bats.

may require something more elaborate than a Gaussian filter to be used when placing synthetic bats.

Another aspect which we are currently testing is if a trained neural network can completely replace the background determination, and subtraction, that is currently used to identify foreground objects. In tracking software this is one of the computationally expensive steps and it is also often one of the most problematic parameters because is not well understood by all users. Further the foreground-background threshold can also vary from frame to frame if the brightness within a video changes.

VI. Summary and Conclusions

Bats are an important study system for many ecological and evolutionary questions, including the benefits of ecosystem services, the impacts of climate change on natural populations, and the potential spread of zoonotic diseases to human populations. The tracking metrics used are straightforward and provide interesting insight into animal behavior. The counting software we are working to supplement is critical for many researchers and conservationists, and it will be freely available on the internet.

Training on synthetic data appears to have been successful in removing bias, working with an 80% accuracy. We believe further refinements in the spacing of the synthesized data will improve that accuracy because currently, it appears that some objects are too far from one another and some are too near.

The result from the training on the combined seven video data set is promising as it reached accuracy values across the board that were greater than 88%, but with values typically a few percent lower than the self-compared values. This leads us to conclude that individual networks corresponding to one video type should be used in particular if a niche group (e.g. the Department of Fish and Wildlife) has a specific video type that will be commonly used as part of an agency-wide protocol, and that a universal network can also be trained for all other video types as a catch-all for anyone else.

We plan to use the techniques discussed in this manuscript to ultimately generate two unbiased networks, one that is specifically trained on all videos used by our collaborators and another that is more universally applicable.

VII. Acknowledgment

We thank the Macaulay Library at the Cornell Lab of Ornithology for use of video recordings of Sandhill Crane (ML319530501) and European Starling (ML412546891).

References

[1] Kunz, T.H., Braun de Torrez, E., Bauer, D., Lobova, T., and Fleming, T.H., Ecosystem services provided by bats. Annuals of the New York Academy of Science, The Year in Ecology and Conservation Biology, Volume 1223, Issue 1, March 2011, https://doi.org/10.1111/j.1749-6632.2011.06004.x

[2] Aziz, S.A, McConkey, K.R., Tanalgo, K., Sritongchuay, T., Low, M.-R., Yong, J.Y., Mildenstein, T. L., Nuevo-Diego, C.E., Lim, V.-C., and Racey, P.A. The critical importance of Old World fruit bats for healthy ecosystems and economies. Frontiers in Ecology and Evolution, Volume 9, April 2021, https://doi.org/10.3389/fevo.2021.641411

[3] Ramirez-Francel, L.A., Garcia-Herrera, L.V., Losada-Prado, S., Reinoso-Florez, G., Sanchez-Hernandez, A., Estrada-Villegas, S., Lim, B.K., and Guevara, G. Bats and their vital ecosystem services: a global review. Integrative Zoology, May 2021, doi: 10.1111/1749-4877.12552

[4] Frick, W.F., Kingston, T., Flanders, J., A review of the major threats and challenges to global bat conservation. Annuals of the New York Academy of Science, The Year in Ecology and Conservation Biology, Volume 1469, Issue 1,June 2020, https://doi.org/10.1111/nyas.14045

[5] Thomson, T.J., Scott, J.A., Castleberry S.B. Evaluation of Methods for Monitoring Long-term Population Trends in Cave-roosting Bats, 2010 Proc. Annu. Conf. South Eastern Association Fish and Wildlife Agencies

[6] Corcoran, A., and Hedrick T. L., ThruTracker Software available on line at: https://sonarjamming.com/thrutracker/

[7] Bentley, I., BatCount Software available on line at: https://sourceforge.net/projects/batcount/

[8] Corcoran, A. J., Schirmacher, M. R., Black, E., and Hedrick, T. L., ThruTracker: Open-Source Software for 2-D and 3-D Animal Video Tracking, BioRxiv, (2021) https://doi.org/10.1101/2021.05.12.443854

[9] Bentley, I. Kuczynska, V., Eddington, V., Armstrong, M. and Kloepper, L. BatCount: A software program to count moving animals. PLoS ONE 18(3): e0278012 (2023).

[10] Mathworks 2023. Object Detection Using Faster R-CNN Deep Learning https://www.mathworks.com/help/deeplearning/ug/object-detection-using-faster-r-cnn-deep-learning.html

[11] Mathworks 2023. Create Simple Deep Learning Neural Network for Classification. https://www.mathworks.com/help/deeplearning/ug/create-simple-deep-learning-network-for-classification.html

[12] Stack Overflow 2023. image processing- background extraction and update from video using matlab. https://stackoverflow.com/questions/24278661/background-extraction-and-update-from-video-using-matlab/41674453#41674453

[13] Hao, D., Zhang, L., Sumkin, J., Mohamed, A., and Wu, S., "Inaccurate Labels in Weakly-Supervised Deep Learning: Automatic Identification and Correction and Their Impact on Classification Performance," in IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 9, pp. 2701-2710, Sept. 2020

[14] Nigam, N., Dutta, T., and Gupta, H.P. (2020). Impact of Noisy Labels in Learning Techniques: A Survey. In: Kolhe, M., Tiwari, S., Trivedi, M., and Mishra, K. (eds) Advances in Data and Information Sciences. Lecture Notes in Networks and Systems, vol 94. Springer, Singapore.

[15] Zhang, C., Bengio, S., Hardt, M., Recht, B., and Vinyals, O., 2016, arXiv, arXiv:1611.03530. doi:10.48550/arXiv.1611.03530

[16] Krawczyk, B. Learning from imbalanced data: open challenges and future directions. Prog Artif Intell 5, 221–232 (2016).

[17] Shorten, C. and Khoshgoftaar, T.M. "A survey on Image Data Augmentation for Deep Learning." Journal of Big Data 6 (2019): 1-48.