Progress toward automated migratory waterfowl census using drones and deep learning

Work-in-Progress

Rowan L. Converse[†]
Department of Geography and Environmental Studies
University of New Mexico
Albuquerque, New Mexico, USA
rowanconverse@unm.edu

Grant M. Harris
Biological Services
US Fish and Wildlife Service
Albuquerque, New Mexico, USA
grant_harris@fws.gov

Matthew J. Butler
Biological Services
US Fish and Wildlife Service
Albuquerque, New Mexico, USA
matthew_butler@fws.gov

Christopher D. Lippitt
Department of Geography and Environmental Studies
University of New Mexico
Albuquerque, New Mexico, USA
clippitt@unm.edu

Steven E. Sesnie
Biological Services
US Fish and Wildlife Service
Albuquerque, New Mexico, USA
steven_sesnie@fws.gov

David R. Stewart
Biological Services
US Fish and Wildlife Service
Albuquerque, New Mexico, USA
david_stewart@fws.gov

ABSTRACT

Wildlife managers invest substantial resources in monitoring populations of migratory waterfowl. Aerial imaging surveys appear to yield more precise counts compared with traditional ground and aerial survey methods; however, given the substantial time investment required to manually interpret aerial imagery of wildlife, automated image interpretation methods, such as deep learning, will be needed to make this approach scalable. We present current progress, continuing problems, and lessons learned from a cooperative research project with the US Fish and Wildlife Service to develop an unoccupied aerial system (UAS) field survey workflow for censusing wintering waterfowl at federally managed wildlife refuges in New Mexico as well as a deployable convolutional neural network (CNN) model for automated detection and classification of waterfowl. Our goal is to develop a scalable workflow that can be deployed at wildlife refuges within the federal system throughout the United States and beyond. Our framework utilizes crowdsourced UAS image annotations from the participatory science platform Zooniverse; we validated these annotations against annotations from wildlife biologists and found that the consensus of the two groups was comparable in enumerating (91%), classifying to general taxonomic group (99.99%), and locating (80%) birds in the imagery. We tested multiple CNN architectures and selected YOLOv5 for its performance. Models trained on the crowdsourced annotations outperform the more limited expert annotations, even when subsampled to the same number of annotations as the expert dataset; thus, our results indicate that suitable training data to finetune a CNN model to a new site may be expeditiously collected with very few UAS transects by focusing on collecting representative variability (i.e., of species, vegetation, environmental conditions, etc.), assuming relatively dense aggregations of target bird species. We plan a 2023 winter field deployment at sites in New Mexico and Texas.

[†]Corresponding author

GeoWildLife '23, November 13, 2023, Hamburg, Germany

ACM ISBN 979-8-4007-0355-3/23/11.
https://doi.org/10.1145/3615893.3628757

CCS CONCEPTS

• Object detection • Object identification • Systems biology

KEYWORDS

UAS, waterfowl, wildlife inventory, participatory science, aerial survey

ACM Reference format:

Rowan L. Converse, Christopher D. Lippitt, Grant M. Harris, Steven E. Sesnie, Matthew J. Butler, and David R. Stewart. 2023. Progress toward automated waterfowl census using drones and deep learning. In *Proceedings of GeoWildlife Workshop, ACM SigSpatial conference (SigSpatial'23). ACM, Hamburg, Germany, 4 pages.* https://doi.org/10.1145/3615893.3628757

1 Introduction

Accurate population counts of migratory waterfowl species are a critical wildlife management tool to ensure appropriate resource availability along migration routes, set harvest levels, and monitor species status (Andersson et al. 2015). Traditional methods for waterfowl census consist of in-time counts, either of total populations or sampling transects, performed on the ground or from low-altitude occupied aircraft (Augustine et al. 2023). Such in-time counts can generate inaccurate indices of animal abundance regardless of the experience level of the observer (Boyd 2000, Frederick et al. 2003)

The increasing affordability of high-resolution sensor systems and unoccupied aerial systems (UAS, "drones") makes an aerial imaging approach for wildlife surveys more feasible than in the recent past. Aerial imaging surveys appear to yield more precise counts of bird populations compared with traditional ground and aerial survey methods (Hodgson et al. 2016) and may offer other benefits including increased safety for participants (Sasse 2003) and less disturbance to wildlife (Chabot et al. 2015). Given the substantial time investment required to manually interpret aerial imagery of wildlife (Hodgson et al. 2016), automated image interpretation methods, such as deep learning, will be needed to make this approach scalable (Lippitt and Zhang 2018).

We present current progress, continuing problems, and lessons learned from a cooperative research project with the US Fish and Wildlife Service to develop a UAS field survey workflow for censusing wintering waterfowl at federally managed wildlife refuges in New Mexico as well as a deployable convolutional neural network (CNN) model for automated detection and classification of general types of waterfowl (Figure 1). The goal of this research program is to develop a scalable workflow that can be deployed at wildlife refuges within the federal system throughout the United States and beyond.

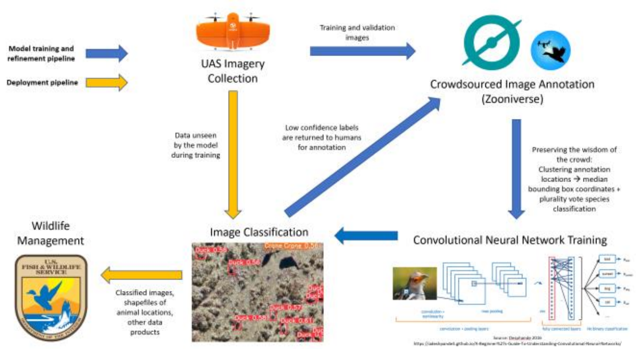


Figure 1: Conceptual workflow for automating waterfowl census with UAS and deep learning with convolutional neural networks, including model training and deployment

2 Methods

UAS flights were conducted at waterfowl management areas in central and northern New Mexico from 2018 - 2023 to collect high-resolution (average 0.87cm/px) imagery of overwintering migratory waterfowl populations. The study area consists of a variety of habitat types: agricultural fields, artificially flooded wetlands, riparian areas, and xeric uplands. Due to the environmental complexity and the small size/cryptic coloring of several of the target species, we consider our application to be a relatively difficult image interpretation case for human observers and thus also a difficult case for deep-learning based object detection and classification (Figure 2).

Figure 2: Example image of waterfowl in an artificially flooded wetland habitat in our study area in central New Mexico, with birds outlined in red boxes.

A twelve-image benchmark set of imagery was annotated to species by fifteen wildlife biologists, while a much larger set of ~30,000 image tiles (corresponding to ~1,000 images) was uploaded to the online participatory science platform Zooniverse to crowdsource identifications of general types of waterfowl (duck/goose/crane). To test the validity of the image labels, we derived a set of consensus annotations from both groups and examined variability within each group both overall and per class. Additionally, we compared the consensus annotations of the two groups to each other to establish the level of comparability between them.

We tested multiple CNN architectures using the expert annotations for training and selected YOLOv5 for its performance; for full results, see Sa'doun et al. 2021. However, the expert-trained models suffered from poor recall due to the limited number of training samples. Therefore, we tested multiple interventions to improve model performance. First, we trained models with the larger set of crowdsourced annotations, and randomly subsampled the dataset to determine if there existed a threshold for model performance improvements so that numeric targets could be set for future data collection efforts. We also tested the impact of transfer learning using a large set of aerial imagery of waterfowl species from around the world used to train a global bird detection model (Weinstein et al. 2022). Finally, we tested the addition of background ("empty") images to the model.

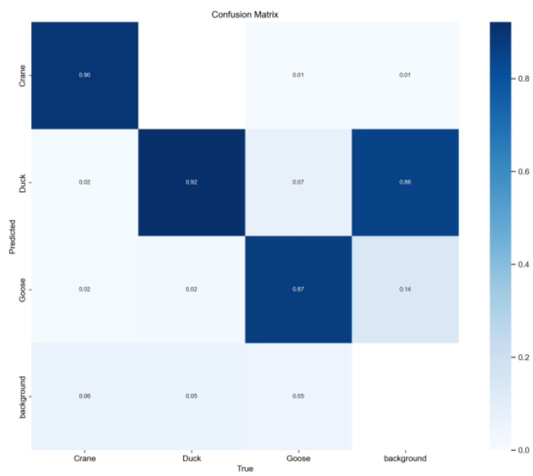
3 Results

3.1 Label Validation

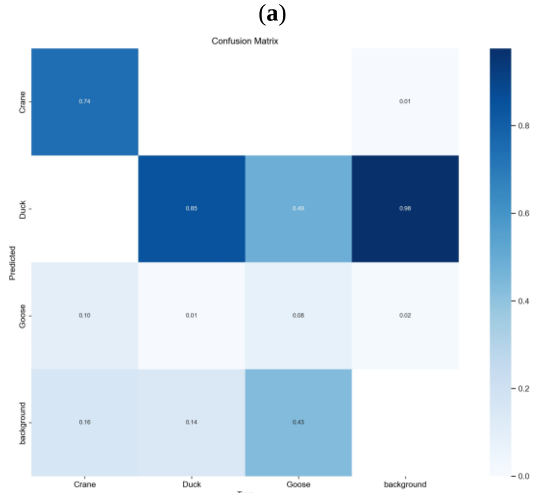
The consensus of the two groups was comparable in enumerating (91%), classifying to morphological groups (99.99%), and locating (80%) birds in the imagery. Within each group, average individual agreement with the group consensus was high for the three classes of interest: duck (0.93 for experts, 0.92 for volunteers), crane (0.99 for experts, 0.95 for volunteers), and goose (0.94 for experts, 0.74 for volunteers). However, we found that experts were highly variable in classifying ducks to species, ranging from an average of 0.83 for the most common duck species class to 0.43 for the rarest.

3.2 Model Validation

Comparing against the same in-domain test set, models trained using the crowdsourced annotations had superior performance compared to models trained only using the expert annotations, even when the crowdsourced dataset was randomly subsampled to a similar size (Figure 3). The addition of background images and the use of transfer learning with the global waterfowl dataset did not significantly improve results.



(a)



(b)

Figure 3: Comparison of classification accuracy of a CNN model trained (a) using 1% of consensus crowdsourced annotations; and (b) the full set of consensus expert annotations

4 Discussion

Based on our label validation study, we believe that volunteers can, in aggregate, produce image annotations of similar quality to experts when classifying broad morphological groups of waterfowl in aerial imagery. Additionally, we found that even expert-generated species-level identifications of ducks in aerial imagery are likely not reliable. It is possible that collecting data at a higher spatial resolution may aid experts in identifying duck species (Dulava et al. 2015). Hierarchical models may also be a feasible method for coping with uncertainty in species classifications (Augustine et al. 2023).

Models trained using the crowdsourced annotations universally outperformed models trained using the expert annotations, even when the crowdsourced dataset was subsampled to a similar size as the expert dataset. We believe this occurred because the crowdsourced dataset contained a more representative variety of species distributions and habitats compared to the expert dataset, which contained more redundant examples (Habib et al. 2019). In terms of field planning for collecting training images at novel sites, our results indicate that suitable training data may be expeditiously collected with as little as a single UAS transect per site (~10-15 images), given that non-empty images will likely contain dozens to hundreds examples of the targets of interest. We recommend that practitioners should focus on collecting representative variety (e.g., of habitat, species, environmental conditions, etc) rather than raw numbers of target examples.

While our best fit model has good recall, it continues to have relatively poor precision with the most common class (duck) due to confusion with the background. As noted previously, the environmental complexity of the study area combined with the natural camouflage of duck species makes this a difficult image interpretation task even for human observers. We are currently testing interventions such as adding in a border class and class weighting to mitigate this issue (Kellenberger et al. 2021).

4.1 Future Directions

We have two immediate goals for future directions of the project: 1) evaluating the impact of the UAS on the movements of animals to determine whether this is a source of bias in population counts derived from UAS imagery; and 2) an expansion of the project area to environments and species assemblages out of the domain of the current model.

While it appears that UAS largely do not provoke overt startle responses in waterfowl when flown in a survey pattern at appropriate flight altitudes (Vas et al. 2015), it is unclear if UAS cause more subtle alterations of bird behavior; e.g. swimming away from the direction of UAS flight. These subtle movements may bias animal population counts if animals move systematically out of view of the UAS sensor. We will examine this issue by comparing spatial distributions and counts of animals in consecutive and non-consecutive overlapped image areas. If these distributions appear to be nonrandom, we will test if they can be related statistically to the direction of UAS flight.

Finally, we are in the planning stages for a field deployment this winter to attempt a comprehensive population survey at sites in New Mexico and Texas. We are currently conducting a power analysis to determine the appropriate UAS survey effort based on test flights in February at our primary field site at Bosque del Apache National Wildlife Refuge. We will also test our own recommendations for training data collection for model refinement developed as part of our model optimization analysis as we expand our study area to the Chenier Plain National Wildlife Refuge Complex in Texas, which contains novel species assemblages unseen by our current model.

ACKNOWLEDGMENTS

Thanks to Hays Barrett, Tyler Eschelman, Mohammad Mustafa Sa'doun, and Jesse Sprague for technical assistance in the field and lab. We also gratefully acknowledge the CV4Ecology Summer School funded by the Resnick Institute for Sustainability at Caltech for support in the development phase of the project.
This work was funded under US Fish and Wildlife Co-operative Agreement Number F17AC0122. The findings and conclusions in this article are those of the author(s) and do not necessarily represent the views of the U.S. Fish and Wildlife Service. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.
The authors have no conflicts of interest to declare.

REFERENCES

[1] Kent Andersson, Craig A. Davis, Grant Harris, and David A. Haukos. 2015. An assessment of non-breeding waterfowl surveys on National Wildlife Refuges in the Central Flyway. *Wildlife Society Bulletin* 39, 1 (2015), 79–86. DOI:https://doi.org/10.1002/wsb.500

[2] Ben C. Augustine, Mark D. Koneff, Bradley A. Pickens, and J. Andrew Royle. 2023. Towards estimating marine wildlife abundance using aerial surveys and deep learning with hierarchical classifications subject to error. 2023.02.20.529272. DOI:https://doi.org/10.1101/2023.02.20.529272

[3] W. Sean Boyd. 2000. A comparison of photo counts versus visual estimates for determining the size of snow goose flocks. *forn* 71, 4 (October 2000), 686–690. DOI:https://doi.org/10.1648/0273-8570-71.4.686

[4] Dominique Chabot, Shawn R. Craik, and David M. Bird. 2015. Population Census of a Large Common Tern Colony with a Small Unmanned Aircraft. *PLOS ONE* 10, 4 (April 2015), e0122588. DOI:https://doi.org/10.1371/journal.pone.0122588

[5] Sharon Dulava, William T. Bean, and Orien M. W. Richmond. 2015. Environmental Reviews and Case Studies: Applications of Unmanned Aircraft Systems (UAS) for Waterbird Surveys. *Environmental Practice* 17, 3 (September 2015), 201–210. DOI:https://doi.org/10.1017/S1466046615000186

[6] Peter C. Frederick, Becky Hylton, Julie A. Heath, and Martin Ruane. 2003. Accuracy and variation in estimates of large numbers of birds by individual observers using an aerial survey simulator. *forn* 74, 3 (July 2003), 281–287. DOI:https://doi.org/10.1648/0273-8570-74.3.281

[7] Ahsan Habib, Chandan Karmakar, and John Yearwood. 2019. Impact of ECG Dataset Diversity on Generalization of CNN Model for Detecting QRS Complex. *IEEE Access* 7, (2019), 93275–93285. DOI:https://doi.org/10.1109/ACCESS.2019.2927726

[8] Jarrod C. Hodgson, Shane M. Baylis, Rowan Mott, Ashley Herrod, and Rohan H. Clarke. 2016. Precision wildlife monitoring using unmanned aerial vehicles. *Scientific Reports* 6, 1 (March 2016). DOI:https://doi.org/10.1038/srep22574

[9] Benjamin Kellenberger, Thor Veen, Eelke Folmer, and Devis Tuia. 2021. 21 000 birds in 4.5 h: efficient large-scale seabird detection with machine learning. *Remote Sensing in Ecology and Conservation* 7, 3 (2021), 445–460. DOI:https://doi.org/10.1002/rse2.200

[10] Christopher D. Lippitt and Su Zhang. 2018. The impact of small unmanned airborne platforms on passive optical remote sensing: a conceptual perspective. *International Journal of Remote Sensing* 39, 15–16 (August 2018), 4852–4868. DOI:https://doi.org/10.1080/01431161.2018.1490504

[11] Mohammad Mustafa Sa'doun, Christopher D. Lippitt, Gernot Paulus, and Karl–Heinrich Anders. 2021. A Comparison of Convolutional Neural Network Architectures for Automated Detection and Identification of Waterfowl in Complex Environments. *giforum* 1, (2021), 152–166. DOI:https://doi.org/10.1553/giscience2021_02_s152

[12] D Blake Sasse. Job-Related Mortality of Wildlife Workers in the United States, 1937-2000. 7.

[13] Elisabeth Vas, Amélie Lescroël, Olivier Duriez, Guillaume Boguszewski, and David Grémillet. 2015. Approaching birds with drones: first experiments and ethical guidelines. *Biology Letters* 11, 2 (February 2015), 20140754. DOI:https://doi.org/10.1098/rsbl.2014.0754

[14] Ben G. Weinstein, Lindsey Garner, Vienna R. Saccomanno, Ashley Steinkraus, Andrew Ortega, Kristen Brush, Glenda Yenni, Ann E. McKellar, Rowan Converse, Christopher D. Lippitt, Alex Wegmann, Nick D. Holmes, Alice J. Edney, Tom Hart, Mark J. Jessopp, Rohan H. Clarke, Dominik Marchowski, Henry Senyondo, Ryan Dotson, Ethan P. White, Peter Frederick, and S. K. Morgan Ernest. 2022. A general deep learning model for bird detection in high-resolution airborne imagery. *Ecological Applications* 32, 8 (2022), e2694. DOI:https://doi.org/10.1002/eap.2694