Automated Sea Duck Counts from Aerial Photographs
Final Report

Prepared for:

US Fish and Wildlife Service
1011 East Tudor Road
Anchorage, Alaska 99503

Prepared by:

Thomas Prebyl and Riley Knoedler

Western EcoSystems Technology, Inc.
415 West 17th Street, Suite 200
Cheyenne, Wyoming 82001

August 5, 2022

WEST

TABLE OF CONTENTS

LIST OF TABLE

LIST OF APPENDICES

INTRODUCTION

Western Ecosystems Technology, Inc. (WEST) is pleased to provide this final report as a deliverable product for Funding Opportunity Number F19AS00298, Sea Duck Joint Venture FY2020 Competitive Grants. In this work we addressed Science Need D, “Develop and/or evaluate methods for efficiently automating counts of birds in aerial photographs of flocks, including birds with varying distributions and density patterns, and uniform vs. dimorphic plumages,” of the funding opportunity. Currently, sea duck population monitoring utilizes aerial surveys and human review of photos to identify and count duck species. As the human review process is labor intensive and time consuming, the primary motivation for this work was to improve efficiency by developing an automated approach to detect sea ducks and other bird species in aerial photos. Specifically, our objective was to develop an artificial intelligence system based on a convoluted neural network capable of detecting and counting sea duck individuals in aerial photos and classifying individuals to species and sex when possible.

METHODS

Image Annotation

We used a dataset consisting of 810 aerial images containing sea ducks and other birds in offshore and coastal environments. Images were collected from fixed wing aircraft at varying flight heights with image resolution ranging from (8688 x 5792 to 10328 x 7760 pixels). A trained biologist reviewed each image identifying and annotating the position of objects of interest (birds) in the image by drawing a bounding box surrounding each object. Each object was assigned to one of the following species categories (King Eider, Steller’s Eider, Spectacled Eider, Common Eider, Long-tailed Duck, Black Scoter, White-winged Scoter, Surf Scoter, Bufflehead, Red-breasted Merganser, Other Duck, Other Bird, Not Bird). The “Other Duck” category consisted of instances when an object was known to be a duck but the species could not reliably be determined. The “Other Bird” category consisted of all non-duck species of birds. Objects were also assigned a sex category (male, female, unknown), when possible. Annotations were reviewed for accuracy and exported in Common Objects in Context (COCO) version 1.0 (Lin et al. 2015) format prior to model training.

Model Training

Prior to training models, we split the dataset into training and validation sets, with 70% used for training and 30% used for validation using the Python library scikit-learn v.0.23.2 (Pedregosa et al. 2011). Due to a limited sample size for many duck species, we pooled species with less than 100 annotated objects into the Other Duck and Other Bird categories. The final species categories used for model training were: Black Scoter, Common Eider, King Eider, Stellar’s Eider, Other Bird, Other Duck, and Not Bird. To improve training efficiency, we resized all images to a width and height of 8688 and 5792 pixels, respectively. We initially explored three types of convoluted neural network model architectures: single shot detectors (Liu et al. 2016), You Only Look Once (Redmon et al. 2016), and Faster R-CNN (Ren et al. 2016). Initial results indicated the Faster R-CNN architecture provided superior performance and we subsequently limited model training efforts to that architecture. Specifically, we used the Faster R-CNN architecture with a ResNet-50 feature pyramid network backbone (Ren et al. 2016) and weights that were pre-trained on the COCO dataset (Lin et al. 2015) from torchvision v.0.9.0 (Marcel and Rodriguez 2010). Anchor box sizes (expected sizes of objects in images) were initially set at 25, 32, 45, 65, and 100 pixels, with aspect ratios of 0.5, 1, and 2. A stochastic gradient descent with momentum for the optimizer was used, with a momentum value of 0.9 and a weight decay rate of 0.0005. We used a plateau-based learning rate optimizer with an initial learning rate of 0.0005 and reduced the learning rate by a factor of 0.005 once the validation loss stopped improving. Validation loss was evaluated in every epoch and retained model weights if the validation loss improved, otherwise the weights were discarded and the best previous weights were loaded in the next epoch. All model training was conducted using the Pytorch v.1.9.1 library (Paszke et al. 2019) in Python v.3.8.12 (Python Software Foundation 2022).

Model Evaluation

We evaluated model performance by deploying the model on the validation dataset (30% of images originally withheld from model training). For each object category in the model, we compared predicted bounding boxes and ground truth annotation for that category and calculated the amount of overlap based on Intersection over Union (IoU) where

$$IoU = \frac{Area\ of\ bounding\ box\ overlap\ between\ ground\ truth\ and\ prediction}{Area\ of\ bounding\ box\ union\ between\ ground\ truth\ and\ prediction}$$

For this study, we defined a true positive detection as a predicted bounding box that overlaps a ground truth annotation with an IoU greater than or equal to 0.10. Many studies set a higher IoU threshold of 0.50. However, for this study we deemed the lower threshold of 0.10 acceptable given the objects (sea ducks) were typically small relative to the camera field of view, thus, an IoU threshold of 0.10 was sufficient to achieve our goals with respect to identifying the position of objects in each image. All predicted bounding boxes that had an IoU less than 0.10 with a ground truth annotation were considered a false positive (FP). Additionally, all ground truth bounding boxes that did not have a matching predicted bounding box with an IoU greater than or equal to 0.10 were considered a false negative (FN). We next calculated recall and precision for each category, where:

$$Recall = \frac{TP}{TP+FN}$$

and

$$Precision = \frac{TP}{TP+FP}$$

The convoluted neural network produces confidence estimates for each object detected that can be used to exclude objects unlikely to belong to any category of interest. In general, excluding objects with low confidence tends to decrease recall but increase precision. To describe this precision-recall tradeoff, we calculated the average precision (AP) for each category, which is a measure of the area under the precision-recall curve following standard methods for evaluating object detection algorithms (Girshick et al. 2016). We then calculated a mean AP across all categories, which summarizes overall model fit. Additionally, we calculated precision and recall at a range of discrete confidence levels (0.05–0.95). Lastly, we calculated a confusion matrix comparing predicted and ground-truth categories when deploying the model using a confidence value that balanced precision and recall (confidence > = 0.10).

RESULTS

The annotation dataset consisted of 36,609 individual birds and 775 other marine species. The number of annotations was not evenly distributed among species categories used in the final model and ranged from 264–16,631 (Table 1). Model performance varied considerably among species categories, image quality, and background sea conditions. The two individual species with the largest training sample sizes (Common Eider and Steller's Eider) had the highest AP values (0.58 and 0.55, respectively; Table 1). AP values were lower for individual species with lower sample sizes (Black Scoter = 0.23 and King Eider = zero). All King Eiders in the validation dataset were predicted to be other species or not detected (Table 2). The two combined categories (Other Duck and Other Bird) also had poor model performance with AP values of 0.08 and 0.45, respectively (Table 1). Despite relatively poor ability to discriminate among species categories, the model successful detected 59% of all birds in the validation dataset (range 42–83% among species categories) when using a confidence threshold of 0.1 and considering an object as detected regardless of the predicted species category (Table 2). Recall and precision rates varied as a function of the confidence threshold used in model deployment, with lower confidence values resulting in higher recall, but lower precision (Appendix A).

Table 1. Model evaluation statistics.

Category	Number of Objects: Training	Number of Objects: Validation	Recall*	Precision[2]	Average Precision
Black Scoter	1,401	458	0.34	0.48	0.23
Common Eider	2,516	1,302	0.61	0.74	0.58
King Eider	53	211	0	0	0
Steller's Eider	2,735	1,626	0.63	0.58	0.55
Other Duck	7,465	2,211	0.18	0.24	0.08
Other Bird	12,931	3,700	0.45	0.35	0.45
Not Bird	429	346	0	0	0
mean	**3,932**	**1,407**	**0.32**	**0.34**	**0.27**

Note: Sample sizes indicated the number of objects (duck, birds) that were present in all training and validation images. Recall and precision rates are based on the validation dataset and deploying the model using a confidence threshold of 0.10.

[1] Recall and precision calculations based on excluding model predictions with a confidence less than 0.1.

Table 2. Confusion matrix comparing predicted object categories and ground truth annotations.

Ground-truth	Predicted: Black Scoter	Predicted: Common Eider	Predicted: King Eider	Predicted: Steller's Eider	Predicted: Other Duck	Predicted: Other Bird	Predicted: Not Bird	Predicted: Empty	Percent Detected[1]
Black Scoter	150	10	0	114	70	6	0	108	76.42
Common Eider	3	765	0	8	5	32	0	489	62.44
King Eider	3	0	0	54	116	2	0	36	82.94

Table 2. Confusion matrix comparing predicted object categories and ground truth annotations.

	Predicted								
Ground-truth	**Black Scoter**	**Common Eider**	**King Eider**	**Steller's Eider**	**Other Duck**	**Other Bird**	**Not Bird**	**Empty**	**Percent Detected[1]**
Steller's Eider	36	0	0	998	152	40	0	400	75.40
Other Duck	60	5	0	351	389	143	0	1263	42.88
Other Bird	27	149	0	23	227	1798	0	1476	60.11
Not Bird	19	0	0	0	35	54	0	238	31.21
Empty	10	17	0	75	307	1382	0	0	–

Note: Values along the diagonal represent true positive detections with an Intersection over Union (IoU) > 0.10. Values in the ground-truth "Empty" row indicate instances where no ground truth annotation overlapped the predicted bounding box with an IoU > 0.10. Values in the predicted "Empty" column indicate instances where a ground-truth annotation existed, but no bounding box overlapped with an IoU > 0.10.

[1] An object was considered detected regardless of the predicted category.

DISCUSSION

The artificial intelligence approach we developed in this work is promising and we were also able to identify key challenges for automated detection of sea ducks in aerial photographs. In certain conditions, the neural network model performed well and detected all ducks and sea birds in a given image without any FP detections. Although we did not specifically quantify sources of FP or FN predictions, we did anecdotally note these error rates varied considerable among images. This high performance was generally observed on photos with calm dark sea conditions, an absence of floating debris, and relatively widely spaced birds. In contrast, the model performed less accurately on images that are more complex. FP detections generally occurred in images where sea foam, floating vegetation, or other bird-sized debris were miss-classified as ducks. In regard to FNs, the model struggles to accurately detect all objects when ducks and birds are in tightly clustered groups. This pattern of FNs is indicative of limited image resolution after images had been resized to accommodate the Faster R-CNN architecture. This model also showed mixed results in terms of correctly distinguishing between bird species. A number of factors likely limited the model's ability to correctly classify species, including 1) insufficient image resolution, 2) poor image quality (motion blur, high flight heights), and 3) limited sample size for some species

Several steps could be taken to improve model performance. The most obvious improvement would be to eliminate image resizing in favor of an image tiling approach to take advantage of the full resolution of the existing imagery. Tiling would entail converting individual images into multiple tiles (e.g., 800 x 800 pixel tiles) rather than reducing individual images to a lower-resolution version compatible with the convoluted neural network architectures. While tiling would decrease processing speed, it is likely to substantially increase model recall and precision. A second clear step that could be taken to improve performance would be to standardize input images for variation in camera settings and flight heights. If metadata on camera focal length and flight heights are available, images could be scaled (or tiled) in such a way that birds in the imagery are a consistent size in terms of pixel dimensions. Additional potential improvements include exploring an object segmentation approach and additional network architectures to better distinguish sea duck species and eliminate FPs associated with floating debris.

REFERENCES

Girshick, R., D. J., T. Darrell, and J. Malik. 2016. Region-Based Convolutional Networks for Accurate Object Detection and Segmentation. IEEE Trans Pattern Anal Mach Intell Jan(38): 142-158. doi: 10.1109/TPAMI.2015.2437384.

Lin, T.-Y., M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, and P. Dollár. 2015. Microsoft COCO: Common Objects in Context. arXiv:1405.0312, Computer Vision and Pattern Recognition, Cornell University, Ithaca, New York. February 21, 2015. Accessed August 2022. Available online: https://arxiv.org/abs/1405.0312

Liu, W., D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg. 2016. SSD: Single Shot Multibox Detector. B. Leibe, J. Matas, N. Sebe, and M. Welling, eds. Computer Vision - European Conference on Computer Vision (ECCV) 2016. ECCV 2016. Lecture Notes in Computer Science. Springer, Cham. Vol. 9905.

Marcel, S. and Y. Rodriguez. 2010. Torchvision the Machine-Vision Package of Torch. Pages 1485–1488 Proceedings of the 18th ACM international conference on Multimedia. Association for Computing Machinery. New York, New York, USA.

Paszke, A., S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala. 2019. Pytorch: An Imperative Style, High-Performance Deep Learning Library. H. Wallach, H. Larochelle, A. Beygelzimer, F. Alché-Buc, E. Fox, and R. Garnett, eds. Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, Vancouver, British Columbia, Canada. December 8-14 2019.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay. 2011. Scikit-Learn: Machine Learning in Python. Journal of Machine Learning Research 12(2011): 2825-2830.

Python Software Foundation (PSF). 2022. Python. PSF, Beaverton, Oregon. Accessed August 2022. Available online: https://www.python.org/

Redmon, J., S. Divvala, R. Girshick, and A. Farhadi. 2016. You Only Look Once: Unified, Real-Time Object Detection. Proceedings of the IEEE conference on computer vision and pattern recognition. 779-788 pp. Available online: https://www.cv-foundation.org/openaccess/content_cvpr_2016/papers/Redmon_You_Only_Look_CVPR_2016_paper.pdf

Ren, S., K. He, R. Girshick, and J. Sun. 2016. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. arXiv:1506.01497, Computer Vision and Pattern Recognition, Cornell University, Ithaca, New York. January 6, 2016. Accessed August 2022. Available online: https://arxiv.org/abs/1506.01497

Appendix A. Recall and Precision Calculations

Appendix A. Recall and precision calculations for each species category at multiple confidence levels.

Category	Recall	Precision	Confidence
Black Scoter	0.34	0.48	0.05
Black Scoter	0.34	0.48	0.10
Black Scoter	0.33	0.48	0.20
Black Scoter	0.32	0.48	0.30
Black Scoter	0.29	0.48	0.40
Black Scoter	0.25	0.48	0.50
Black Scoter	0.19	0.49	0.60
Black Scoter	0.12	0.45	0.70
Black Scoter	0.08	0.51	0.80
Black Scoter	0.02	0.71	0.90
Black Scoter	0	0	0.95
Common Eider	0.63	0.74	0.05
Common Eider	0.61	0.74	0.10
Common Eider	0.59	0.74	0.20
Common Eider	0.55	0.74	0.30
Common Eider	0.49	0.73	0.40
Common Eider	0.42	0.73	0.50
Common Eider	0.36	0.72	0.60
Common Eider	0.28	0.71	0.70
Common Eider	0.21	0.70	0.80
Common Eider	0.09	0.68	0.90
Common Eider	0.03	0.73	0.95
King Eider	0	0	0.05
King Eider	0	0	0.10
King Eider	0	0	0.20
King Eider	0	0	0.30
King Eider	0	0	0.40
King Eider	0	0	0.50
King Eider	0	0	0.60
King Eider	0	0	0.70
King Eider	0	0	0.80
King Eider	0	0	0.90
King Eider	0	0	0.95
Steller's Eider	0.64	0.58	0.05
Steller's Eider	0.63	0.58	0.10
Steller's Eider	0.61	0.58	0.20
Steller's Eider	0.59	0.58	0.30
Steller's Eider	0.55	0.59	0.40
Steller's Eider	0.48	0.60	0.50
Steller's Eider	0.41	0.61	0.60
Steller's Eider	0.33	0.62	0.70
Steller's Eider	0.24	0.63	0.80
Steller's Eider	0.12	0.68	0.90
Steller's Eider	0.02	0.82	0.95
Other Duck	0.21	0.24	0.05
Other Duck	0.18	0.24	0.10
Other Duck	0.16	0.24	0.20
Other Duck	0.14	0.24	0.30
Other Duck	0.11	0.23	0.40
Other Duck	0.08	0.25	0.50
Other Duck	0.03	0.20	0.60
Other Duck	0.01	0.15	0.70
Other Duck	0	0.08	0.80
Other Duck	0	0	0.90
Other Duck	0	0	0.95
Other Bird	0.53	0.34	0.05

Appendix A. Recall and precision calculations for each species category at multiple confidence levels.

Category	Recall	Precision	Confidence
Other Bird	0.45	0.35	0.10
Other Bird	0.38	0.35	0.20
Other Bird	0.33	0.35	0.30
Other Bird	0.28	0.35	0.40
Other Bird	0.25	0.37	0.50
Other Bird	0.23	0.39	0.60
Other Bird	0.20	0.44	0.70
Other Bird	0.18	0.51	0.80
Other Bird	0.16	0.59	0.90
Other Bird	0.14	0.63	0.95
Not Bird	0	0	0.05
Not Bird	0	0	0.10
Not Bird	0	0	0.20
Not Bird	0	0	0.30
Not Bird	0	0	0.40
Not Bird	0	0	0.50
Not Bird	0	0	0.60
Not Bird	0	0	0.70
Not Bird	0	0	0.80
Not Bird	0	0	0.90
Not Bird	0	0	0.95