A convolutional neural network for automated detection of humpback whale song in a diverse, long-term passive acoustic dataset: Supplementary Material

Authors
Ann N. Allen[1]*, Matt Harvey[2], Lauren Harrell[3], Aren Jansen[2], Karlina P. Merkens[4], Carrie C. Wall[5,6], Julie Cattiau[2], Erin M. Oleson[1]

[1]NOAA Pacific Islands Fisheries Science Center, Honolulu HI, United States.
[2]Google, Inc., Mountain View, CA, United States.
[3]Google, Inc., Boulder, CO, United States.
[4]Lynker Technologies, LLC, Portland, OR. Under contract to NOAA Pacific Islands Fisheries Science Center, Honolulu, HI, United States.
[5]University of Colorado Boulder, Boulder, CO, United States.
[6]NOAA National Centers for Environmental Information, Boulder, CO, United States.

*Corresponding author: ann.n.allen@gmail.com

| Site	Initial	Model-dependent 1	Model-dependent 2	Segments	Model-dependent 3	Model-Dependent 4	Validation	Expansion of Test Set	Total	Fraction of Annotations from Site	Fraction of Audio from Site
Kona	2733	455	1,260	650	79		904	217	7,361	24.20%	29.40%
		(261)	(672)		(11)			(119)			
Cross SM		11	1,098	137	85	289	319	106	2,224	7.30%	0.90%
		(2)	(79)		(19)	(12)		(67)			
Kauai	1458	70	981	39	61		2,670	2	5,507	18.10%	5.90%
		(2)	(53)		(18)			(71)			
Pearl and Hermes Reef	333	24	330	23	15	274	180	280	2,460	8.10%	8.80%
		(14)	(564)		(72)	(26)		(325)			
Ladd SM			179		5		170	8	1,207	4.00%	1.30%
			(593)		(187)			(65)			
Kingman Reef	22		276		4			2	1,074	3.50%	1.70%
			(618)		(81)			(71)			
Palmyra			2		(93)		202	(205)	535	1.80%	5.40%
			(33)								
Equator			4		97			(70)	496	1.60%	1.40%
			(325)								
Wake		21	179	20	1	185	297	68	4,006	13.20%	4.30%
			(2,625)		(109)	(126)		(375)			
Pagan								132	349	1.10%	3.10%
								(217)			
Tinian		122	30	149	4	60	326	149	2,044	6.70%	17.20%
			(563)		(101)	(245)		(295)			
Saipan		12	1,220	53	65	265	289	147	3,102	10.20%	20.60%
		(6)	(617)		(34)	(37)		(357)			
Total	4546	1,000	12,301	1,071	1,141	1,519	5,357	3,430	30,365		

SUPPLEMENTARY TABLE 1. Count of annotations by site and audit effort. Parenthesized values are counts of explicit negatives. In the columns without parenthesized values, all regions of audio that are in the sample but not marked as positive comprise the implicit negatives for the effort in that column.

|	Training	Validation (Model Metrics)	Test (Plots & Tables)
Duration Audited	66.7 hours	80.0 hours	145.1 hours
Duration Marked Positive	5.45 hours	1.9 hours	27.4 hours
Duration Marked Explicitly Negative	3.1 hours	0 hours	45.5 hours
Duration of Implicit Negatives	57.2 hours	74.1 hours	117.7 hours
Number of Positive Annotations	13,140	5,357	6,600
Number of Explicit Negative Annotations	9,179	0	2,187
Number of Segments with Any Positive	6,773	195	1,365
Number of Segments with No Positive	9,365	3,452	5,617
Amount Audited as Percent of Corpus	0.04%	0.04%	0.08%

SUPPLEMENTARY TABLE 2. Scale of the annotation effort. The total audio duration of the entire dataset is over 20 years; only a very small fraction of that was audited.

|		Average Precision	AUC-ROC
Compression Method	PCEN	0.90	0.99
	log	0.84	0.96
	root	0.77	0.95
Context Window Length (s)	1.28	0.94	1.00
	3.84	0.97	1.00
	6.4	0.96	0.99

SUPPLEMENTARY TABLE 3. Compression and time scale experiment metrics computed on the validation set from the checkpoint with highest average precision.
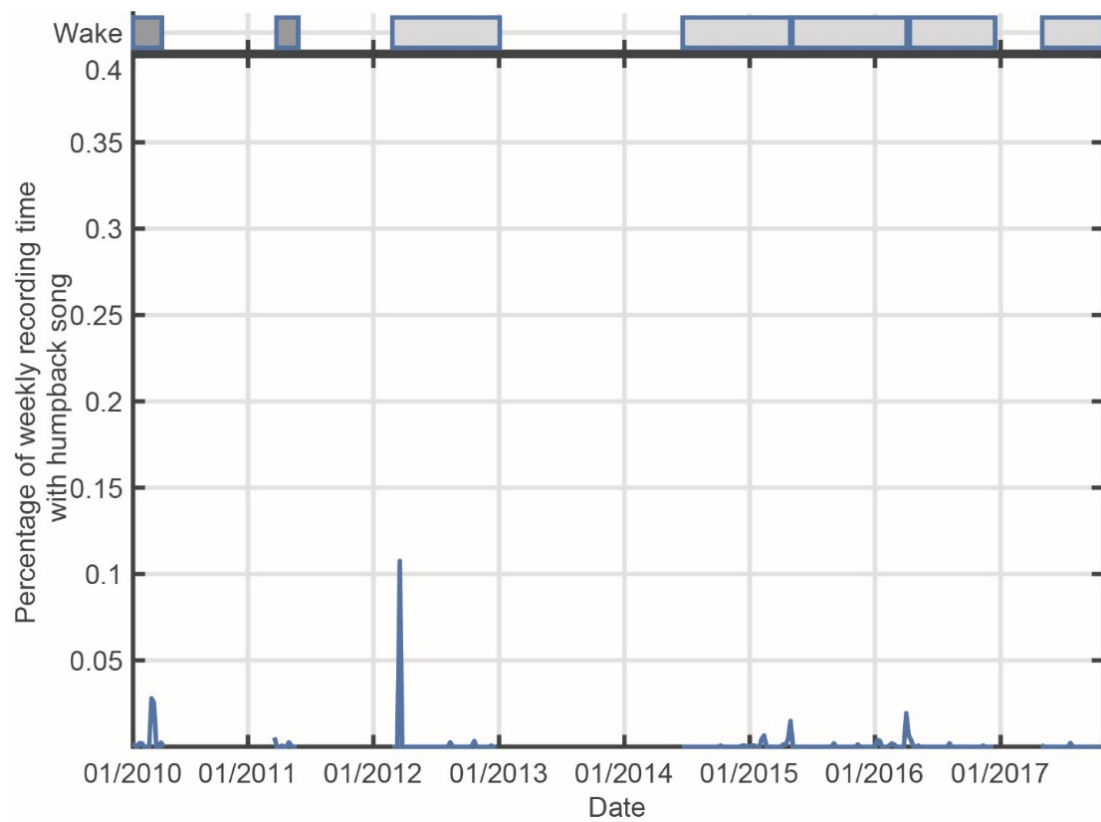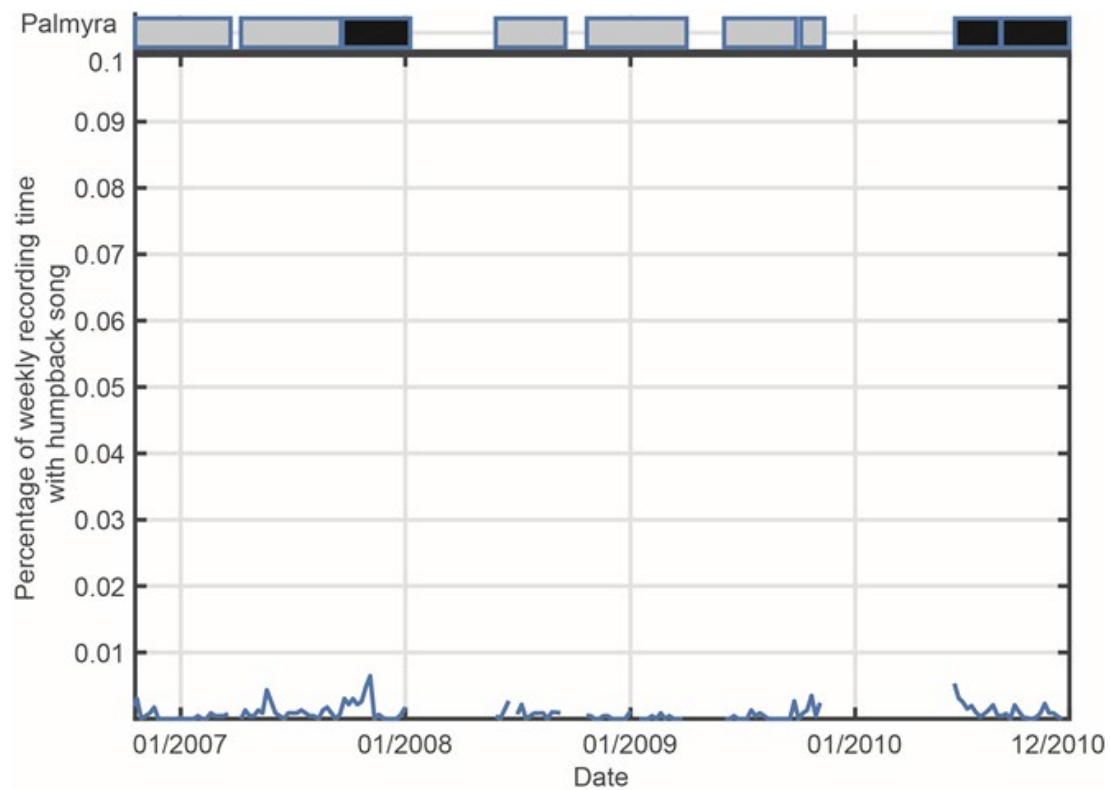
Supplementary Figure 1. Percentage of recording time per week containing positive detections at Wake Island. Shaded boxes represent the time periods that have recordings, with shading indicating duty cycle, which ranged from 16.7% to 50% recording time.
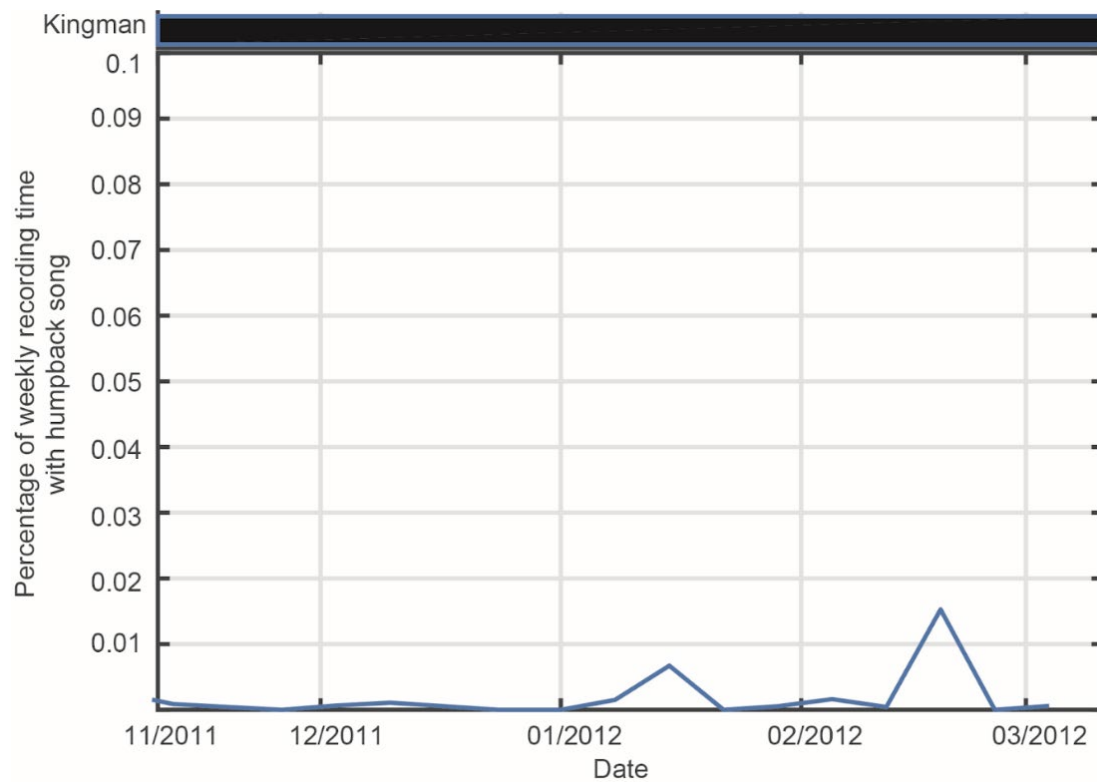
Supplementary Figure 2. Percentage of recording time per week containing positive detections at Kingman Reef. Recording duty cycle was continuous for Kingman Reef.
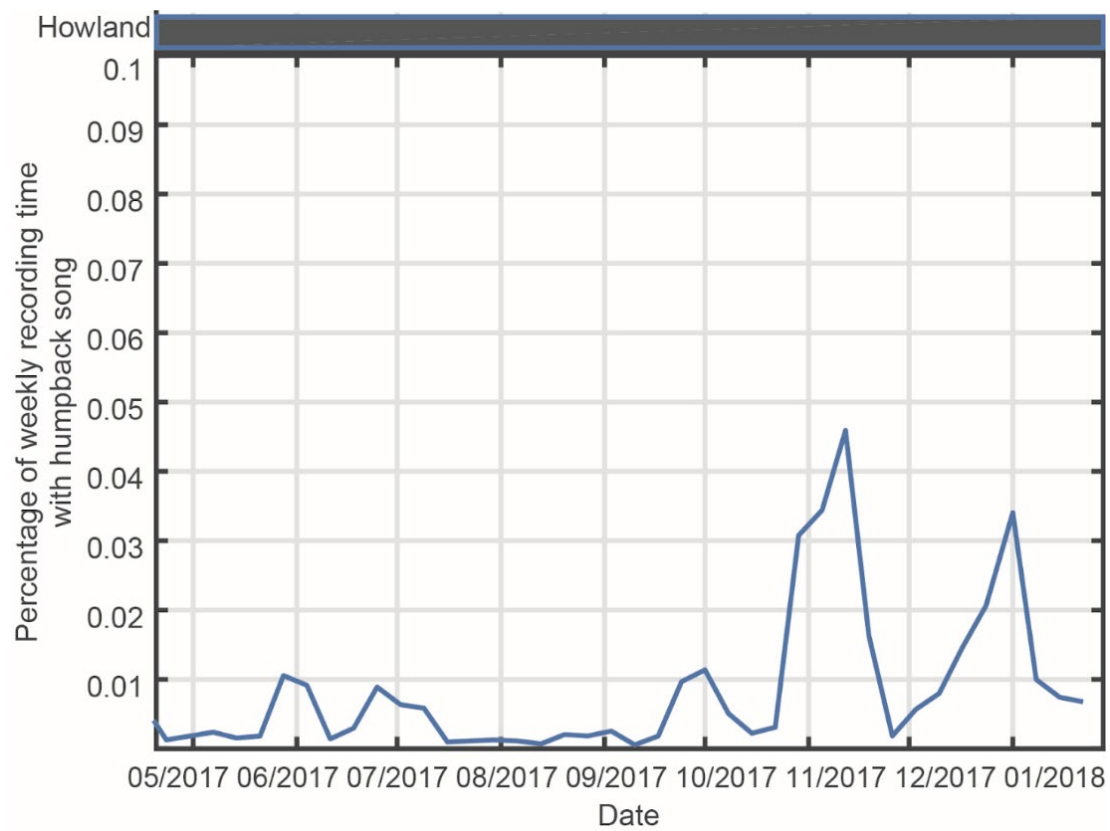
Supplementary Figure 3. Percentage of recording time per week containing positive detections at Howland Reef. Duty cycle was 80% for Howland Island.

Supplementary Figure 4. Percentage of recording time per week containing positive detections at Palmyra Island. Shaded boxes represent the time periods that have recordings, with shading indicating duty cycle, which ranged from 25% to continuous recording time.