Supplementary Information for

A deep active learning system for species identification and counting in camera trap images

Mohammad Sadegh Norouzzadeh, Dan Morris, Sara Beery, Neel Joshi, Nebojsa Jojic, and Jeff Clune

Methods in Ecology and Evolution 2020

September 15, 2020

Corresponding Author: Dan Morris (dan@microsoft.com)

This PDF file includes:
Supplementary text
Figures S.1 to S.2
Tables S.1 to S.4
SI References Cited

S.1 Triplet loss

The triplet loss was originally designed for problems with a variable number of classes, such as human face recognition (Schroff et al., 2015). Recent studies (Hermans et al., 2017) showed the effectiveness of the triplet loss in learning a useful encoding. The triplet loss tries to put samples with the same label nearby in the embedding space and samples with different labels far away in the embedding space. To train a network with the triplet loss, we arrange the labeled examples into triplets. Each triplet consists of a baseline sampled image (the anchor), another sampled image with the same class as the anchor (positive), and a sampled image belonging to a different class (negative). For a distance metric d and a triplet (A, P, N), the triplet loss (which optimization attempts to minimize) is defined as:

$$L = \max(d(A, P) - d(A, N) + \text{margin}, 0) \tag{1}$$

In Eq. 1, *margin* is a hyperparameter specifying that $d(A, N)$ must be at least *margin* greater than $d(A, P)$.

During training, we select N samples from the dataset, where N is the batch size. To do so, we first select K classes uniformly at random from all possible classes, then select P samples per class randomly from the dataset. We then create all possible triplets within this set of N samples, by taking each of the N samples in turn, making it the anchor, and creating all possible triplets for it. The values of N, K, and P are hyperparameters of the algorithm that we set differently for different datasets (see our code implementation for the specific values per dataset).

According to the above definition of the triplet loss (Eq. 1), there are two types of samples: (1) *satisfied triplets*, which have a loss of zero because they already satisfy the condition of the triplet loss; i.e., the negative sample is more than *margin* further from the anchor than the positive sample is to the anchor (2) *unsatisfied triplets*, where the loss is positive. Because satisfied triplets have a loss of zero they have no effect on training the weights of the network. Therefore, we omit them from training. Various strategies could be utilized to select triplets such as choosing the hardest negative (the unsatisfied triplet with maximal loss) or randomly choosing unsatisfied triplets. We follow the original triplet loss paper (Schroff et al., 2015) by randomly selecting unsatisfied triplets during training.

S.2 Active learning selection criteria

Many query selection criteria have been proposed in the literature; for our experiments, we employ two criteria based on model uncertainty (confidence-based and margin-based selection (Settles and Craven, 2008)) and three criteria based on identifying dense regions in the input space (informative diverse (Dasgupta and Hsu, 2008), margin cluster mean (Xu et al., 2003), and k-Center (Sener and Savarese, 2017)). In this section, we summarize each of these criteria. For more details on active learning query selection criteria, refer to (Settles and Craven, 2008).

S.2.1 Model uncertainty selection

Both the confidence-based and margin-based techniques belong to the model uncertainty selection category. The main assumption of these approaches is that when the underlying model is uncertain about predicting a sample, that sample could be more informative than the others. The uncertainty measure is interpreted from the model's output.

The confidence-based approach chooses the samples for which the model has the lowest confidence in the most probable class; the margin-based approach chooses the samples with the smallest gap between the model's most confident and second-most confident classes.

S.2.2 Density-based selection

The primary assumption of criteria that select based on density is that, for learning efficiently, we should not only query the labels of uncertain samples, but should also query those samples that are representative of many inputs, i.e. *dense* regions of the underlying input space. This assumption makes density-based methods less likely to select outliers, and thus more informative about most of the data. The informative diverse technique (Dasgupta and Hsu, 2008) first forms a hierarchical clustering of the unlabeled samples and then selects active learning queries so that the distribution of queries matches the distribution of the entire data. The margin cluster mean criterion (Xu et al., 2003) clusters the samples lying within the margin of an SVM classifier trained on the labeled samples, and then selects the samples at cluster centers. The k-center method (Sener and Savarese, 2017), which has the best performance in our experiments, chooses a set of samples such that a model trained over the selected subset performs equally well on the remaining samples. The k-center method achieves this goal by defining the problem of active learning as a core-set selection problem (Agarwal et al., 2005) and then solving it.

References

Pankaj K Agarwal, Sariel Har-Peled, and Kasturi R Varadarajan. Geometric approximation via coresets. *Combinatorial and computational geometry*, 52:1–30, 2005.

Sanjoy Dasgupta and Daniel Hsu. Hierarchical sampling for active learning. In *Proceedings of the 25th international conference on Machine learning*, pages 208–215. ACM, 2008.

Alexander Hermans, Lucas Beyer, and Bastian Leibe. In defense of the triplet loss for person re-identification. *arXiv preprint arXiv:1703.07737*, 2017.

Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: a unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 815–823, 2015.

Ozan Sener and Silvio Savarese. Active learning for convolutional neural networks: a core-set approach. *arXiv preprint arXiv:1708.00489*, 2017.

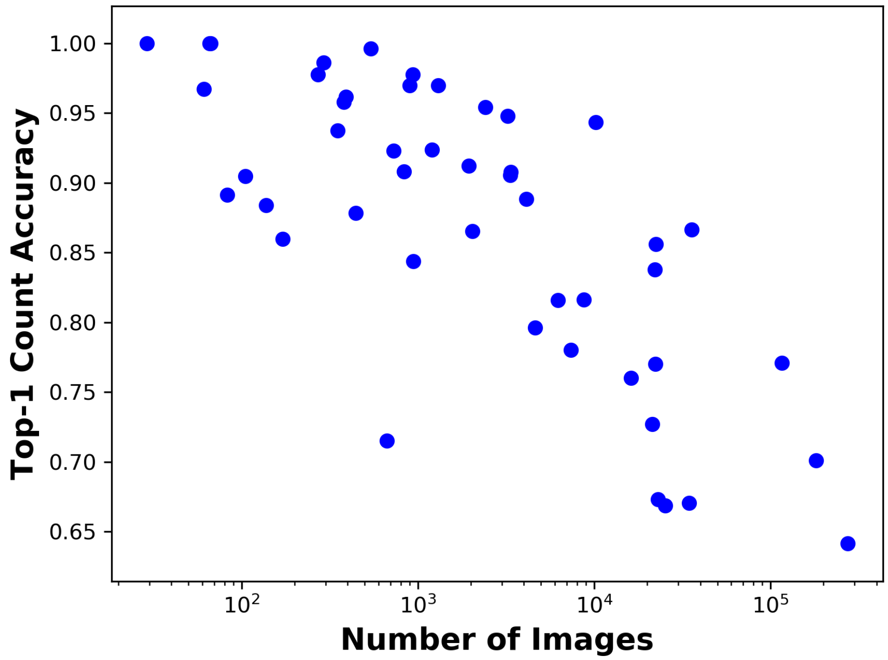

Figure S.1: The number of images in the dataset (which was not used to train the pretrained model) per species vs. the accuracy of the pretrained model in its ability to predict the number of animals in an image containing that species. Interestingly, the model's ability to count accurately decreases with more common animals (for which there are more images in the test set). One possible explanation is that more common animals (e.g. zebra, wildebeest) often appear together in larger numbers, making the task more difficult (e.g. declaring there to be one rhinoceros vs. two is easier than declaring whether there are 8, 9, or 10. These data are presented in tabular form in Table S.1.

Burr Settles and Mark Craven. An analysis of active learning strategies for sequence labeling tasks. In *Proceedings of the conference on empirical methods in natural language processing*, pages 1070–1079. Association for Computational Linguistics, 2008.

Zhao Xu, Kai Yu, Volker Tresp, Xiaowei Xu, and Jizhi Wang. Representative sampling for text classification using support vector machines. In *European Conference on Information Retrieval*, pages 393–407. Springer, 2003.

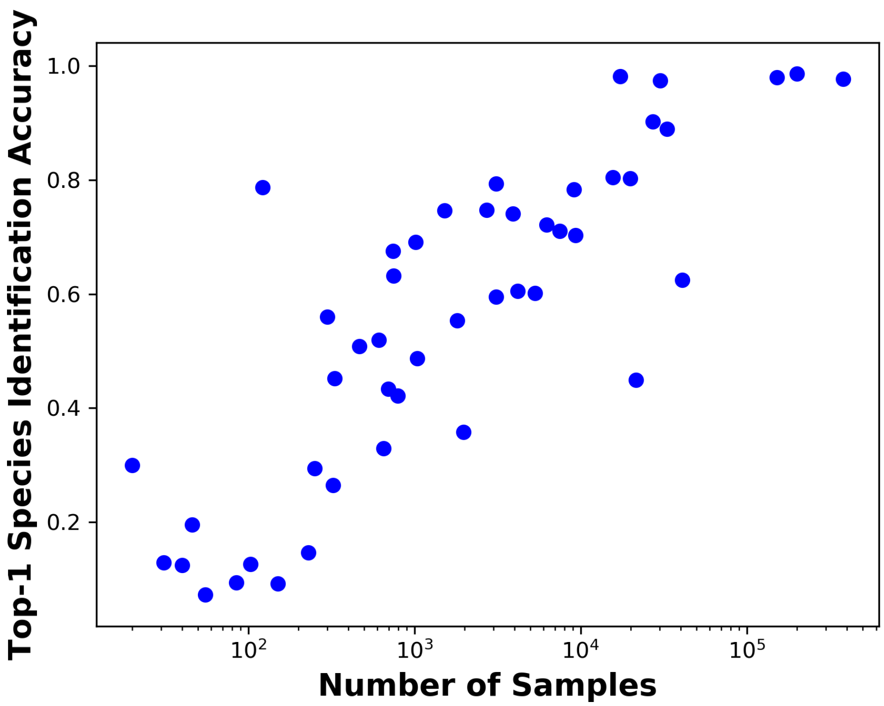

Figure S.2: Number of crops per species vs. identification accuracy for each species. This figure suggests even in case of active learning, accuracy is correlated with the number of crops in the overall dataset for a species. That suggests that active learning does not itself solve the problem of data imbalance, wherein performance tends to be higher for overrepresented data classes and poor for rare classes. This phenomenon likely occurs because there are more chances for crops from common species to satisfy the loss function (e.g. cause disagreement). These data are presented in tabular form in Table S.2.

Table S.1: The accuracy of counting for each species in the Snapshot Serengeti dataset. *Number of Images* indicates the total number of images containing that species in the overall dataset (but recall that counts are generated by a pretrained model, not a model trained on these images). *Correct Count* indicates the number of images that where counted correctly by the detector model. Interestingly, the model's ability to count accurately decreases with more common animals (for which there are more images in the test set). One possible explanation is that more common animals (e.g. zebra, wildebeest) often appear together in larger numbers, making the task more difficult (e.g. declaring there to be one rhinoceros vs. two is easier than declaring whether there are 8, 9, or 10. The information in this table is visualized in Fig. S.1.

Species	**Number of Images**	**Correct Count**	**Accuracy**
Aardvark	542	540	0.996
Aardwolf	292	288	0.986
Baboon	4,618	3,677	0.796
BatEaredFox	729	673	0.923
Buffalo	34,685	23,255	0.670
Bushbuck	352	330	0.938
Caracal	171	147	0.860
Cheetah	3,354	3,037	0.905
Civet	67	67	1.000
DikDik	3364	3054	0.908
Eland	7,395	5,770	0.780
Elephant	25,294	16,908	0.668
Gazelle Grants	21,344	15,517	0.727
Gazelle Thomsons	116,442	89,765	0.771
Genet	61	59	0.967
Giraffe	22,439	19,205	0.856
Guinea Fowl	23,024	15,496	0.673
Hare	900	873	0.970
Hartebeest	35,669	30,906	0.866
Hippopotamus	3,231	3,063	0.948
Honey Badger	83	74	0.892
HyenaSpotted	10,242	9,664	0.944
HyenaStriped	271	265	0.978
Impala	22,281	17,156	0.770
Jackal	1,207	1,115	0.924
KoriBustard	2,042	1,767	0.865
Leopard	382	366	0.958
Lion Female	8,773	7,161	0.816
Lion Male	2,413	2,303	0.954
Mongoose	670	479	0.715
Ostrich	1,945	1,774	0.912
Other Bird	16,240	12,344	0.760
Porcupine	444	390	0.878
Reedbuck	4,131	3,670	0.888
Reptiles	391	376	0.962
Rhinoceros	66	66	1.000
Rodents	138	122	0.884
SecretaryBird	1,302	1,263	0.970
Serval	936	915	0.978
Topi	6,247	5,098	0.816
Vervet Monkey	940	793	0.844
Warthog	22,050	18,471	0.838
Waterbuck	837	760	0.908
Wildcat	105	95	0.905
Wildebeest	275,081	176,409	0.641
Zebra	181,400	127,153	0.701
Zorilla	29	29	1.000

Table S.2: The identification accuracy for each species in the Snapshot Serengeti test set. The model is the margin-based active learning model at the end of 30,000 active learning queries. *Number of Samples* indicates the total number of crops containing that species. *Correct Classifications* indicates the number of crops that where identified correctly by the classification model. The information in this table is visualized in Fig. S.2.

Species	**Number of Samples**	**Correct Classifications**	**Accuracy**
Aardvark	468	238	0.509
Aardwolf	251	74	0.295
Baboon	3,924	2,907	0.741
Bateared Fox	653	215	0.329
Buffalo	40,860	25,509	0.624
Bushbuck	332	150	0.452
Caracal	151	14	0.093
Cheetah	3101	2460	0.793
Civet	46	9	0.196
Dikdik	3,111	1,850	0.595
Eland	7,484	5,318	0.711
Elephant	15,607	12,559	0.805
Gazelle Grants	21,554	9,685	0.449
Gazelle Thomsons	152,265	149,183	0.980
Genet	31	4	0.129
Giraffe	17,365	17,037	0.981
Guineafowl	30,096	29,322	0.974
Hare	745	503	0.675
Hartebeest	33,200	29,510	0.889
Hippopotamus	2,714	2,027	0.747
Honeybadger	55	4	0.073
Hyena Spotted	9,121	7,139	0.783
Hyena Striped	231	34	0.147
Impala	27,172	24,507	0.902
Jackal	1,039	506	0.487
Koribustard	1,811	1,003	0.554
Leopard	325	86	0.265
Lion Female	9,300	6,535	0.703
Lion Male	1,976	708	0.358
Mongoose	610	317	0.520
Ostrich	1,516	1,132	0.747
Other Bird	5,338	3,211	0.602
Porcupine	300	168	0.560
Reedbuck	4,181	2,530	0.605
Reptiles	122	96	0.787
Rhinoceros	40	5	0.125
Rodents	103	13	0.126
Secretary Bird	1,020	705	0.691
Serval	797	336	0.422
Topi	6,246	4,506	0.721
Vervet Monkey	752	475	0.632
Warthog	19,876	15,946	0.802
Waterbuck	699	303	0.433
Wild Cat	85	8	0.094
Wildebeest	381,516	372,741	0.977
Zebra	200,085	197,283	0.986
Zorilla	20	6	0.300

Table S.3: The identification accuracy for each species in the NACTI dataset. The model is the margin-based active learning model at the end of 30,000 active learning queries. Number of samples indicates the total number of crops containing that species. Correct classification indicates the number of crops that where identified correctly by the classification model.

Species	**Number of Crops**	**Correct Classification**	**Accuracy**
American Black Bear	27,452	16910	0.616
American Marten	1,218	250	0.205
American Red Squirrel	2,517	1,239	0.492
Black-tailed Jackrabbit	1,071	896	0.837
Bobcat	24,272	19,934	0.821
California Ground Squirrel	26,780	22,742	0.849
California Quail	2,948	1,904	0.646
Cougar	14,374	10,907	0.759
Coyote	20,452	15,068	0.737
Domestic Cow	2,804,020	2,743,624	0.978
Domestic Dog	1,167	176	0.151
Donkey	3,025	141	0.047
Eastern Gray Squirrel	24,253	19884	0.820
Elk	21,982	16,331	0.743
European Badger	143	1	0.007
Gray Fox	9,828	8,424	0.857
Gray Jay	73	1	0.014
Horse	139	1	0.007
Moose	10,703	6,809	0.636
Mule Deer	94,964	74,072	0.780
Nine-banded Armadillo	7,401	6,172	0.834
North American Porcupine	462	30	0.065
North American River Otter	599	18	0.030
Raccoon	31,357	24,944	0.795
Red Deer	240,576	211,694	0.880
Red Fox	1584	499	0.315
Snowshoe Hare	12,818	10,525	0.821
Striped Skunk	10,389	9,283	0.894
Unidentified Accipitrid	270	3	0.011
Unidentified Bird	77,493	54,582	0.704
Unidentified Chipmunk	884	107	0.121
Unidentified Corvus	1,464	420	0.287
Unidentified Deer	103,411	89,858	0.869
Unidentified Deer Mouse	96	5	0.052
Unidentified Pack Rat	594	281	0.473
Unidentified Rabbit	4,252	2,784	0.655
Virginia Opossum	1,088	596	0.548
Wild Boar	138,182	102,765	0.744
Wild Turkey	4,638	1,179	0.254
Wolf	497	161	0.324
Yellow-bellied Marmot	232	13	0.056

Table S.4: The identification accuracy for each species in the standard test set and location-withheld test set of the Snapshot Serengeti dataset. The model is the margin-based active learning model at the end of 30,000 active learning queries. *Number of samples* indicates the total number of crops containing that species in the test set. *Correct classifications* indicates the number of crops that were identified correctly by the classification model.

Species	**Number of crops in standard test set**	**Correct classifications**	**Accuracy**	**Number of crops in R***	**Correct classifications**	**Accuracy**
Aardvark	472	298	0.631	4	3	0.750
Aardwolf	235	116	0.494	14	7	0.500
Baboon	3,951	2,870	0.726	2	2	1.000
BatEared Fox	648	191	0.295	13	4	0.308
Buffalo	40,865	27,697	0.677	623	408	0.655
Bushbuck	333	95	0.285	0	0	N/A
Caracal	148	46	0.311	1	0	0.000
Cheetah	3018	2206	0.731	79	76	0.962
Civet	49	1	0.020	0	0	N/A
DikDik	3,123	1,875	0.600	22	0	0.000
Eland	5,748	3,972	0.691	193	92	0.477
Elephant	15,376	11,023	0.717	239	186	0.778
Gazelle Grants	20,803	10,664	0.512	1,389	722	0.520
Gazelle Thomsons	148,197	138,612	0.935	4,686	4,180	0.892
Genet	32	0	0.000	0	0	N/A
Giraffe	16,864	15,319	0.908	392	372	0.949
Guinea Fowl	29,507	27,618	0.936	220	194	0.882
Hare	719	507	0.705	11	2	0.182
Hartebeest	32,276	26,542	0.822	1,024	842	0.822
Hippopotamus	2,747	2,048	0.746	0	0	N/A
Honey Badger	51	0	0.000	0	0	N/A
Hyena Spotted	8,769	6,599	0.753	311	222	0.714
Hyena Striped	235	38	0.162	9	3	0.333
Impala	28,004	24,022	0.857	7	0	0.000
Jackal	984	392	0.398	10	3	0.300
KoriBustard	1,749	1,009	0.577	26	10	0.385
Leopard	287	55	0.192	0	0	N/A
Lion Female	9,430	5,917	0.627	19	3	0.158
Lion Male	2,004	617	0.308	8	1	0.125
Mongoose	564	264	0.468	60	43	0.717
Ostrich	1,457	811	0.557	58	30	0.517
Other Bird	5,211	2,751	0.527	98	28	0.286
Porcupine	305	125	0.410	0	0	N/A
Reedbuck	4,262	2,669	0.626	29	4	0.138
Reptiles	132	50	0.379	0	0	N/A
Rhinoceros	43	2	0.047	0	0	N/A
Rodents	103	12	0.117	0	0	N/A
Secretary Bird	960	605	0.630	60	44	0.733
Serval	796	386	0.485	3	2	0.667
Topi	6,150	3,887	0.632	150	89	0.593
Vervet Monkey	755	519	0.687	0	0	N/A
Warthog	19,558	13,990	0.715	367	242	0.659
Waterbuck	715	245	0.343	3	0	0.000
Wildcat	87	7	0.080	2	0	0.000
Wildebeest	357,237	337,875	0.946	10,096	9,563	0.947
Zebra	188,702	184,220	0.976	4,536	4,478	0.987
Zorilla	25	0	0.000	0	0	N/A