The iNaturalist Species Classification and Detection Dataset - Supplementary Material

Grant Van Horn[1] Oisin Mac Aodha[1] Yang Song[2] Yin Cui[3] Chen Sun[2]
Alex Shepard[4] Hartwig Adam[2] Pietro Perona[1] Serge Belongie[3]

[1]Caltech [2]Google [3]Cornell Tech [4]iNaturalist

1. Additional Classification Results

We performed an experiment to understand if there was any relationship between real world animal size and prediction accuracy. Using existing records for bird [4] and mammal [2] body sizes we assigned a mass to each of the classes in iNat2017 that overlapped with these datasets. For a given species, mass will vary due to the life stage or gender of the particular individual. Here, we simply take the average value. This resulted in data for 795 species, from the small Allen's hummingbird (*Selasphorus sasin*) to the large Humpback whale *Megaptera novaeangliae*. In Fig. 1 we can see that median accuracy decreases as the mass of the species increases. These results are preliminary, but reinforce the observation that it can be challenging for humans to take good photographs of larger mammals. More analysis of these failure cases may allow us to produce better, species-specific, instructions for the photographers on iNaturalist.

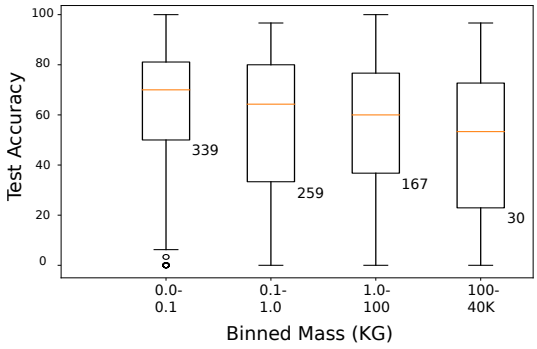

Figure 1. Top one public test set accuracy per class for [6] for a subset of 795 classes of birds and mammals binned according to mass. The number of classes appears to the bottom right of each box.

The IUCN Red List of Vulnerable Species monitors and evaluates the extinction risk of thousands of species and subspecies [1]. In Fig. 2 we plot the Red List status of 1,568 species from the iNat2017 dataset. We see that the vast majority of the species are in the 'Least Concern' category and that test accuracy decreases as the threatened status increases. This can perhaps be explained by the reduced number of images for these species in the dataset.

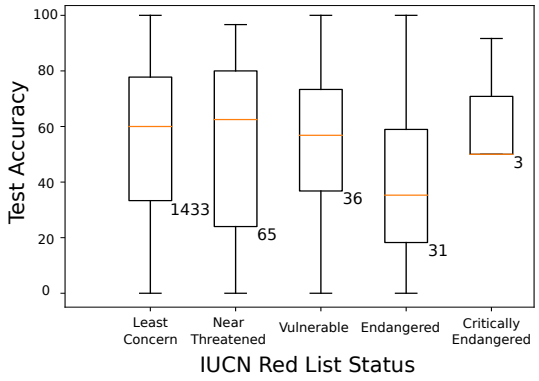

Figure 2. Top one public test set accuracy for [6] for a subset of 1,568 species binned according to their IUCN Red List of Threatened Species status [1]. The number of classes appears to the bottom right of each box.

Finally, in Fig. 3 we examine the relationship between the number of images and the validation accuracy. The median number of training images per class for our entire training set is 41. For this experiment, we capped the maximum number of training images per class to 10, 20, 50, or all, and trained a separate Inception V3 for each case. This corresponds to starting with 50,000 for the case of 10 images per class and then doubling the total amount of training data each time. For each species, we randomly selected the images up until the maximum amount. As noted in the main paper, more attention is needed to improve performance in the low data regime.

1.1. iNat2017 Competition Results

From April to mid July 2017, we ran a public challenge on the machine learning competition platform Kaggle[1] us-

[1]www.kaggle.com/c/inaturalist-challenge-at-fgvc-2017

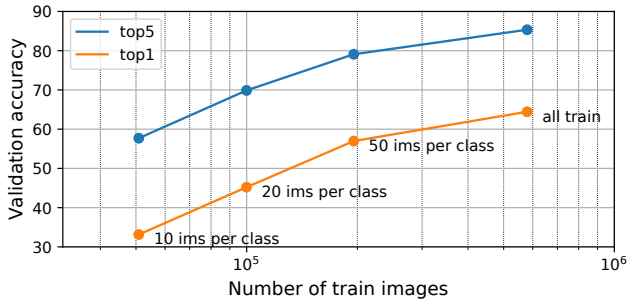

Figure 3. As the maximum number of training images per class increases so does the accuracy. However, we observe diminishing returns as the number of images increases. Results are plotted on the validation set for the Inception V3 network [7].

ing iNat2017. Similar to the classification tasks in [5], we used the top five accuracy metric to rank competitors. We used this metric as some species can only be disambiguated with additional data provided by the observer, such as location or date. Additionally, in a small number of cases multiple species may appear in the same image (*e.g.* a bee on a flower). Overall, there were 32 submissions and we display the final results for the top five teams in Table 1.

The top performing entry from *GMV* consisted of an ensemble of Inception V4 and Inception ResNet V2 networks [6]. Each model was first initialized on the ImageNet-1K dataset and then finetuned with the iNat2017 training set along with 90% of the validation set, utilizing data augmentation at training time. The remaining 10% of the validation set was used for evaluation. To compensate for the imbalanced training data, the models were further fine-tuned on the 90% subset of the validation data that has a more balanced distribution. To address small object size in the dataset, inference was performed on 560×560 resolution images using twelve crops per image at test time.

The additional training data amounts to 15% of the original training set, which along with the ensembling, multiple test crops, and higher resolution account for the improved 81.58% top 1 public accuracy compared to our best performing single model which achieved 68.53%.

Rank	Team name	Public Test		Private Test	
		Top1	Top5	Top1	Top5
1	GMV	**81.58**	**95.19**	**81.28**	**95.13**
2	Terry	77.18	93.60	76.76	93.50
3	Not hotdog	77.04	93.13	76.56	93.01
4	UncleCat	77.64	93.06	77.44	92.97
5	DLUT_VLG	76.75	93.04	76.19	92.96

Table 1. Final public challenge leaderboard results. 'Rank' indicates the final position of the team out of 32 competitors. These results are typically ensemble models, trained with higher input resolution, with the validation set as additional training data.

2. Additional Detection Results

In Table 2 we investigate detector performance for the 2,854-class model across different bounding box sizes using the size conventions of the COCO dataset [3]. As expected, performance is directly correlated with size, where smaller objects are more difficult to detect. However, examining Table 3 we can see that total number of these small instances is low for most super-classes.

	AP^S	AP^M	AP^L	AR^S	AR^M	AR^L
Insecta	13.4	34.7	51.8	13.5	38.9	67.7
Aves	11.5	41.7	55.1	13.3	49.2	69.9
Reptilia	0.0	12.4	22.0	0.0	16.3	46.5
Mammalia	6.7	27.8	37.1	9.0	36.1	55.8
Amphibia	0.0	23.2	29.9	0.0	28.7	54.9
Mollusca	17.5	30.8	35.8	17.5	33.6	55.9
Animalia	24.0	22.7	37.1	26.7	28.2	52.0
Arachnida	16.2	32.9	46.5	16.2	38.5	61.6
Actinopterygii	5.0	16.3	36.1	5.0	17.9	51.1
Overall	11.0	34.7	46.7	12.5	40.7	63.7

Table 2. Super-class level Average Precision (AP) and Average Recall (AR) with respect to object sizes. S, M and, L denote small (area $< 32^2$), medium ($32^2 \leq$ area $\leq 96^2$) and, large (area $> 96^2$) objects. The AP for each super-class is calculated by averaging the results for all species belonging to it. Best and worst performance for each metric are marked by green and red, respectively.

	Small	Medium	Large
Insecta	445	2432	16429
Aves	2375	8898	16239
Reptilia	32	400	5426
Mammalia	280	1068	2751
Amphibia	20	253	2172
Mollusca	74	466	1709
Animalia	72	414	1404
Arachnida	12	152	909
Actinopterygii	32	144	634

Table 3. The number of super-class instances at each bounding box size in the validation set. While AP and AR is low for some super-classes at a particular size (see Table 2), the actual number of instances at that size may also be low.

References

[1] J. Baillie, C. Hilton-Taylor, and S. N. Stuart. *2004 IUCN red list of threatened species: a global species assessment*. IUCN, 2004. 1

[2] K. E. Jones, J. Bielby, M. Cardillo, S. A. Fritz, J. O'Dell, C. D. L. Orme, K. Safi, W. Sechrest, E. H. Boakes, C. Carbone, et al. Pantheria: a species-level database of life history, ecology, and geography of extant and recently extinct mammals. *Ecology*, 2009. 1

[3] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick. Microsoft COCO: Common objects in context. In *ECCV*, 2014. 2

[4] T. Lislevand, J. Figuerola, and T. Székely. Avian body sizes in relation to fecundity, mating system, display behavior, and resource sharing. *Ecology*, 2007. 1

[5] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al. Imagenet large scale visual recognition challenge. *IJCV*, 2015. 2

[6] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. *arXiv preprint arXiv:1602.07261*, 2016. 1, 2

[7] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the inception architecture for computer vision. In *CVPR*, 2016. 2