Transfer learning with fewer ImageNet classes

Michal Kucer, PhD; Diane Oyen, PhD. | ISR-3 / CCS-3, Los Alamos National Laboratory, Los Alamos, New Mexico

Introduction

- ImageNet pre-trained models used for various downstream tasks such as image retrieval and video classification
- ImageNet pre-training (as opposed to pre-training on larger datasets) is computationally feasible on a modest GPU
- **Can we improve representations learned on ImageNet?**

Hard classes

- Evaluate validation accuracy of **ImageNet1000** on all 1000 classes
- Classes with lowest validation accuracy are "**hard**"
- Remove 100 or 200 hardest classes, and re-train to create **ImageNet900** and **ImageNet800**

Q1) Should removing hard classes improve the accuracy of ImageNet models?

- Curriculum learning [1] leads us to postulate that the answer is ***yes***

Q2) Would improving the accuracy on a subset of images improve the transfer learning performance on downstream tasks?

- "Do better ImageNet models transfer better?" [2], leads us to postulate ***yes***
- However, removing classes gives up the volume of data and diversity of images, so it is an open question

Experiments and results

ImageNet experiments

- We explore the effect of removing "hard" classes on the validation accuracy of the 800 common classes
- Validation accuracy **decreases** as more classes are **added** to the training set
- Additional classes are likely being confused with some of the original classes

Transfer learning experiments

- We explore the performance of these representations on 12 downstream tasks in both fixed-feature and full-network transfer [3].
- We see that transfer learning is sometimes **improved** by training with **fewer** classes
- Robustifying the model can remove the advantage of having more classes

Number of classes	$\epsilon = 0.$	$\epsilon = 1.$	$\epsilon = 3.$
800	78.27 / 92.88	71.44 / 88.81 I 56.34 / 81.34	60.70 / 80.86 I 37.50 / 64.87
900	77.42 / 92.56	69.97 / 87.87 I 55.25 / 80.09	60.35 / 81.29 I 36.47 / 63.02
1000	76.23 / 91.80	68.65 / 87.01 I 54.05 / 79.05	59.30 / 80.37 I 35.61 / 61.78

Table 1. Table showing the validation accuracies (Top-1 / Top-5 accuracy) of the various models on the ImageNet dataset on the subset of 800 classes with the best validation accuracy. The left column of the table shows the number of classes that were used to train the ImageNet model. Columns 3 and 4 show the validation accuracies of l2-robust ImageNet models - left pair shows the accuracy on the clean ImageNet validation set, and right pair shows the accuracy on the adversarial validation set with ε equal to the robustness of the particular model.

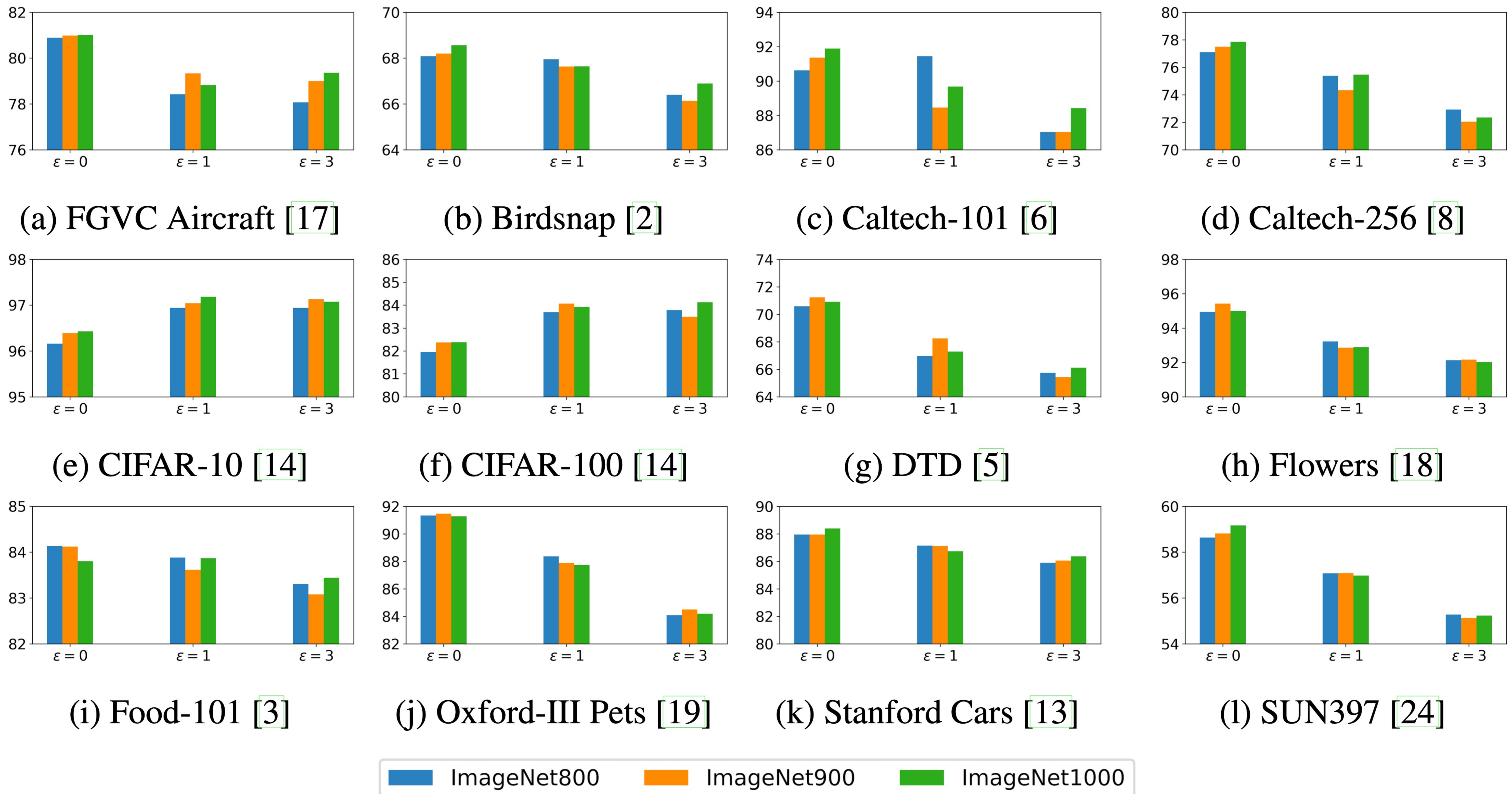

Figure 1. Graphs comparing the full-network transfer learning accuracy across 12 datasets at different levels of robustness and number of ImageNet classes used for training of the baseline models.

References

[1 Bengio, Y., Louradour, J., Collobert, R., & Weston, J. (2009, June). Curriculum learning. In *Proceedings of the 26th annual international conference on machine learning* (pp. 41-48).

[2] Kornblith, S., Shlens, J., & Le, Q. V. (2019). Do better imagenet models transfer better?. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 2661-2671).

[3] Salman, Hadi, et al. "Do adversarially robust imagenet models transfer better?." *arXiv preprint arXiv:2007.08489* (2020).