ImageNet-21K Pretraining for the Masses - Rebuttal Experiments

1 Impact of Different Number of Training Samples

In Figure 1 and Figure 2 we test the impact of the number of training samples in ImageNet-21K on upstream and downstream results.

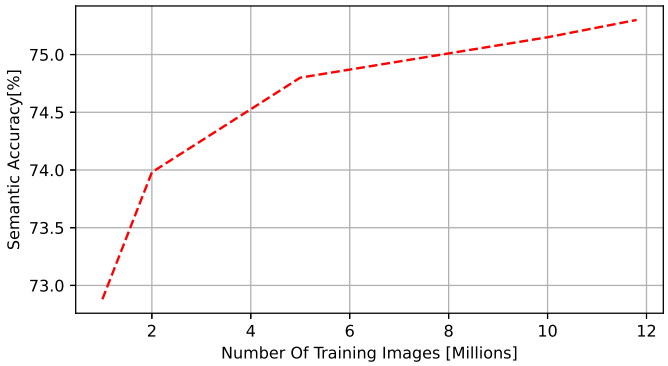

Figure 1: Upstream results for different number of training images.

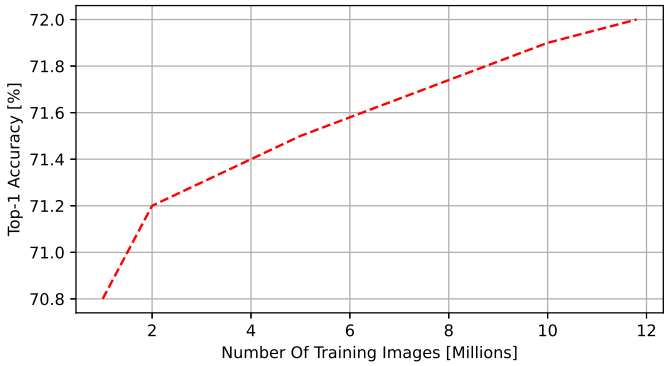

Figure 2: Downstream results for different number of training images, on Inaturalist Dataset.

2 Pretraining Comparisons on Non-classification Tasks

In Table 1 and Table 2 we compare 1K and 21K pretraining on two additional computer-vision tasks: object detection (MS-COCO) and image retrieval (INRIA holidays).

2.1 Object Detection

	1K Pretraining	21K Pretraining
mAP [%]	42.9	44.3

Table 1: Comparing downstream results on MS-COCO object detection dataset.

2.2 Image Retrieval

	1K Pretraining	21K Pretraining
mAP [%]	81.1	82.1

Table 2: Comparing downstream results on on INRIA Holidays image retrieval dataset.

3 Impact of Pretraining on Large Downstream Datasets

In Table 3 we compare downstream results on Open Images datasets, once when using ImageNet-21K pretraining, and once when doing random initialization.

	No Pretraining	21K Pretraining
mAP [%]	80.3	86.0

Table 3: Comparing downstream results for Open Images dataset.

4 Comparison to Other Large-scale Datasets Pretraining

In Table 4 we compare downstream results when using two different datasets for pretraining: ImageNet-21K (semantic softmax training) and Open Images (multi-label training).

Dataset	ImageNet-21K Pretrain	Open Images Pretrain
ImageNet1K[(1)]	**81.4**	81.0
iNaturalist[(1)]	**72.0**	70.7
Food 251[(1)]	**75.8**	74.8
CIFAR 100[(1)]	**90.4**	89.4
MS-COCO[(2)]	**81.3**	80.5
Pascal-VOC[(2)]	**89.7**	89.6
Kinetics 200[(3)]	**83.0**	81.6

Table 4: Comparing ImageNet-21K pretraining to Open Images pretraining. Downstream dataset types and metrics: (1) - single-label, top-1 Acc.[%] ; (2) - multi-label, mAP [%]; (3) - action recognition, top-1 Acc. [%].