Supplementary Materials for Meta-Learning for Semi-Supervised Few-Shot Classification

Mengye Ren **Eleni Triantafillou*** **Sachin Ravi*** **Jake Snell** **Kevin Swersky**

Joshua B. Tenenbaum **Hugo Larochelle** **Richard S. Zemel**

A *tiered*Imagenet Dataset Details

Each high-level category in *tiered*ImageNet contains between 10 and 30 ILSVRC-12 classes (17.8 on average). In the ImageNet hierarchy, some classes have multiple parent nodes. Therefore, classes belonging to more than one category were removed from the dataset to ensure separation between training and test categories. Test categories were chosen to reflect various levels of separation between training and test classes. Some test categories (such as "working dog") are fairly similar to training categories, whereas others (such as "geological formation") are quite different. The list of categories is shown below and statistics of the dataset can be found in Table 1. A visualization of the categories according to the ImageNet hierarchy is shown in Figure 1. The full list of classes per category will also be made public, however for the sake of brevity we do not include it here.

Table 1: Statistics of the *tiered*ImageNet dataset.

	Train	Val	Test	Total
Categories	20	6	8	34
Classes	351	97	160	608
Images	448,695	124,261	206,209	779,165

Train Categories: n02087551 (hound, hound dog), n02092468 (terrier), n02120997 (feline, felid), n02370806 (ungulate, hoofed mammal), n02469914 (primate), n01726692 (snake, serpent, ophidian), n01674216 (saurian), n01524359 (passerine, passeriform bird), n01844917 (aquatic bird), n04081844 (restraint, constraint), n03574816 (instrument), n03800933 (musical instrument, instrument), n03125870 (craft), n04451818 (tool), n03414162 (game equipment), n03278248 (electronic equipment), n03419014 (garment), n03297735 (establishment), n02913152 (building, edifice), n04014297 (protective covering, protective cover, protection).

Validation Categories: n02098550 (sporting dog, gun dog), n03257877 (durables, durable goods, consumer durables), n03405265 (furnishing), n03699975 (machine), n03738472 (mechanism), n03791235 (motor vehicle, automotive vehicle),

Test Categories: n02103406 (working dog), n01473806 (aquatic vertebrate), n02159955 (insect), n04531098 (vessel), n03839993 (obstruction, obstructor, obstructer, impediment, impedimenta), n09287968 (geological formation, formation), n00020090 (substance), n15046900 (solid).

B Extra Experimental Results

Figure 2 shows test accuracy values with different number of unlabeled items during test time. We observe clear improvement in test accuracy when the number grows from 0 to 25. Note that our models were trained with $M = 5$ and thus are showing an ability to extrapolate in generalization.

*These authors contributed equally.

31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

This confirms that, through meta-training, the models learned to acquire a better representation that will be more helpful after semi-supervised refinement. Figure 3 shows our mask output value distribution of the masked soft k-means model on Omniglot. The mask values have a bi-modal distribution, corresponding to distractor and non-distractor items.

C Hyperparameter Details

For Omniglot, we adopted the best hyperparameter settings found for ordinary Prototypical Networks in [1]. In these settings, the learning rate was set to 1e-3, and cut in half every 2K updates starting at update 2K. We trained for a total of 20K updates. For *mini*Imagenet and *tiered*ImageNet, we trained with a starting learning rate of 1e-3, which we also decayed. We started the decay after 25K updates, and every 25K updates thereafter we cut it in half. We trained for a total of 200K updates. We used ADAM [2] for the optimization of our models. For the MLP used in the Masked Soft k-Means model, we use a single hidden layer with 20 hidden units with a tanh non-linearity for all 3 datasets. We did not tune the hyparameters of this MLP so better performance may be attained with a more rigorous hyperparameter search.

References

[1] Jake Snell, Kevin Swersky, and Richard S. Zemel. Prototypical networks for few-shot learning. In *Advances in Neural Information Processing Systems 30*, 2017.

[2] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

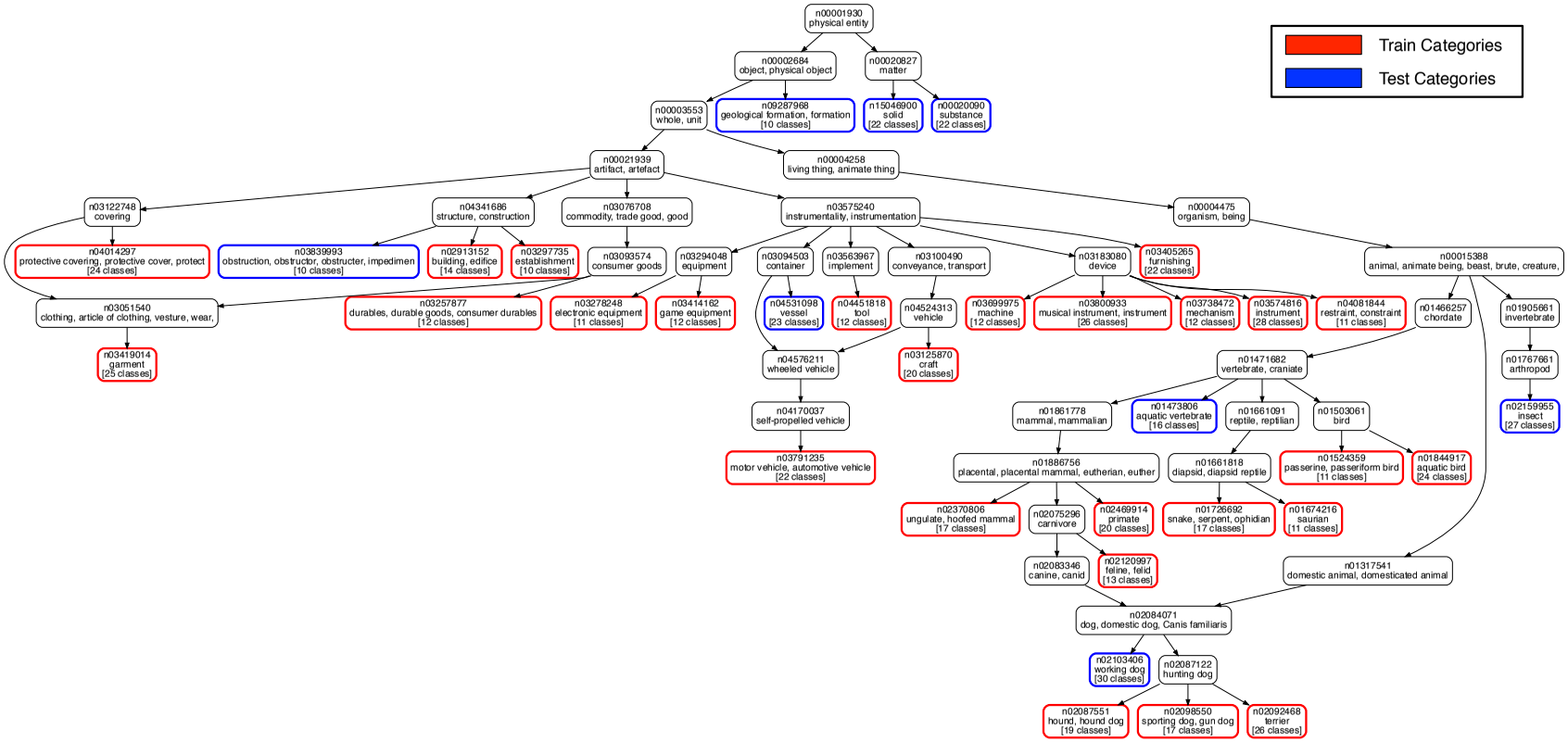

Figure 1: Hierarchy of *tiered*Imagenet categories. Training categories are highlighted in red and test categories in blue. Each category indicates the number of associated classes from ILSVRC-12. Best viewed zoomed-in on electronic version.

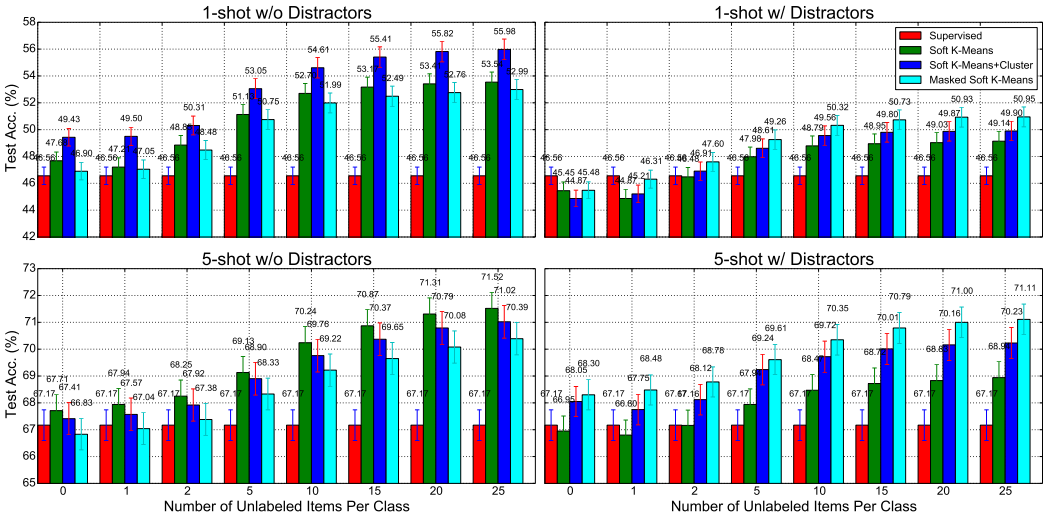

Figure 2: Model Performance on *tiered*ImageNet with different number of unlabeled items during test time. We include test accuracy numbers in this chart.

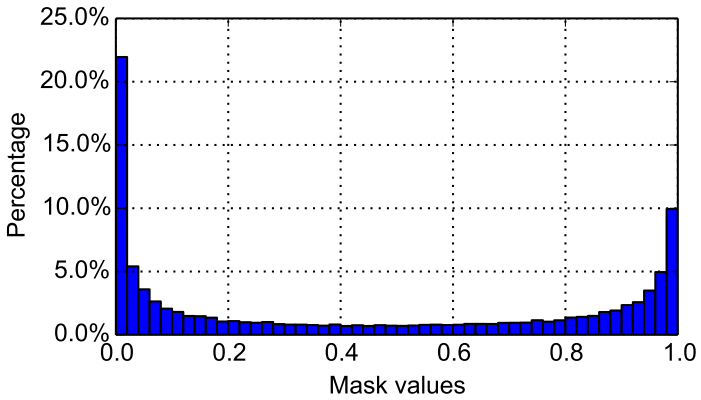

Figure 3: Mask values predicted by masked soft k-means on Omniglot.