Supplemental Material for "Improving Few-Shot Visual Classification with Unlabelled Examples"

Anonymous authors
Paper under double-blind review

A Benchmarks and Training

A.1 Meta-Dataset

A brief description of the sampling procedure used in the Meta-Dataset setting is already provided in Section 4.1. This sampling procedure, however, comes with additional specifications that are uniform across all tasks (such as count enforcing) and dataset specific details such as considering the class hierarchy in ImageNet tasks. The full algorithm for sampling is outlined in (Triantafillou et al., 2019), and we refer the interested reader to Section 3.2 in (Triantafillou et al., 2019) for complete details. This procedure results in a task distribution where most tasks have fewer than 10 classes and each class has fewer than 20 support examples. The task frequency relative to the number of classes is presented in Figure 1a, and the class frequency as compared to the class shot is presented in Figure 1b. The query set contains between 1 and 10 (inclusive) examples per class for all tasks; fewer than 10 query examples occur only when there are not enough total images to support 10 query examples.

A.2 mini/tiered-ImageNet

Task sampling across both mini-ImageNet and tiered-ImageNet first starts by defining a constant number of ways and shots that will be used for each generated task. For a L-shot K-way problem setting, first K classes are sampled from the dataset with uniform probability. Then, for each sampled class, L of the class images are sampled with uniform probability and used as the support examples for the class. In addition, 10 query images (distinct from the support images) are sampled per class.

A.3 Meta-Dataset Training/Testing

Following Bateni et al. (2020) and Requeima et al. (2019), we train our ResNet18 feature extractor as a supervised multi-class image classifier on the training split of the ImageNet subset of the Meta-

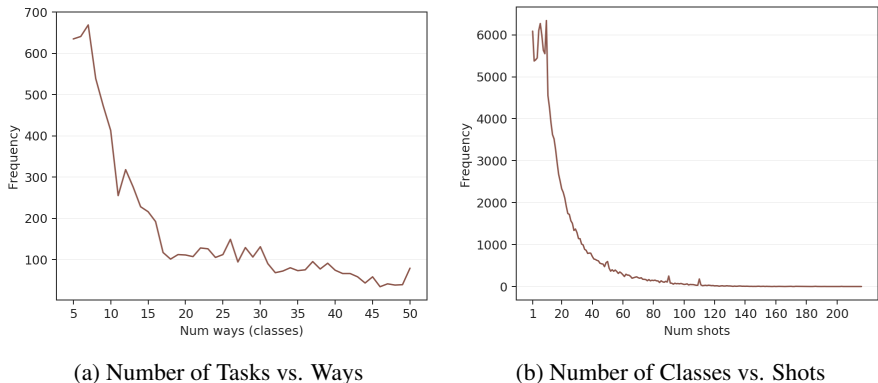(a) Number of Tasks vs. Ways

(b) Number of Classes vs. Shots

Figure 1: Test-time way and shot frequency graphs. Figure is directly from (Bateni et al., 2020). As shown, most tasks have fewer than 10 classes (way) and most classes have less than 20 support examples (shot).

Dataset. We directly follow the procedure described in C.1.1 of Requeima et al. (2019) and use their released checkpoint. Images from the 712 ImageNet classes designated for training by Meta-Dataset (Triantafillou et al., 2019) are first resized to 84x84. The ResNet18 in Transductive CNAPS is then trained as a 712-class image recognition task using this data. Training is done using the cross-entropy loss for 125 epochs using SGD with momentum of 0.9, weight decay of 0.0001, batch size of 256, and a learning rate of 0.1 that is reduced by a factor of 10 every 25 epochs. During training, the dataset was augmented with random crops, horizontal flips, and color jitter.

Once the ResNet18 is trained, we freeze the parameters and proceed to train the adaptation network using Episodic training (Snell et al., 2017; Finn et al., 2017) where tasks themselves are used as training examples. For each iteration of Episodic training, a task (with additional ground truth query labels) is generated, and the adaptation network is trained to minimize classification error (cross entropy) of the query set given the task. We train for a total of 110K tasks, with 16 tasks per batch, resulting in 6875 gradient updates. We train using Adam optimizer with learning rate of 5×10^{-4}. We evaluate on the validation splits of all 8 in-domain and 1 out-of-domain (MSCOCO) datasets, saving the best performing checkpoint for test-time evaluation.

A.4 MINI/TIERED-IMAGENET TRAINING/TESTING

Similar to Meta-Dataset, we first train the ResNet18 feature extractor. This is done with respect to two settings: first, we directly use the training data from mini-ImageNet and tiered-ImageNet, with 38,400 and 448,695 images respectively. Second, we consider a larger training set of 825 classes from ImageNet Russakovsky et al. (2015) that don't overlap with the test sets of the benchmarks. In both cases, we follow the same procedure as the one described for Meta-Dataset in A.3 with the exception of training for 90 epochs and reducing the learning rate every 30 epochs.

After training the ResNet18, the weights are frozen while we train the adaptation network using Episodic training: at each iteration, a task is generated, and we backpropagate the query set classification loss through the adaptation network. For mini/tiered-ImageNet, we train for a total of 20K tasks, validating performance every 2K tasks and saving the best checkpoint for test-time evaluation. We, similarly, use the Adam optimiser with learning rate of 5×10^{-4}, and use a batch size of 16, for a total of 1250 gradient steps.

REFERENCES

Peyman Bateni, Raghav Goyal, Vaden Masrani, Frank Wood, and Leonid Sigal. Improved few-shot visual classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pp. 1126–1135. JMLR. org, 2017.

James Requeima, Jonathan Gordon, John Bronskill, Sebastian Nowozin, and Richard E Turner. Fast and flexible multi-task classification using conditional neural adaptive processes. *arXiv preprint arXiv:1906.07697*, 2019.

Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3):211–252, 2015.

Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. In *Advances in Neural Information Processing Systems*, pp. 4077–4087, 2017.

Eleni Triantafillou, Tyler Zhu, Vincent Dumoulin, Pascal Lamblin, Kelvin Xu, Ross Goroshin, Carles Gelada, Kevin Swersky, Pierre-Antoine Manzagol, and Hugo Larochelle. Meta-dataset: A dataset of datasets for learning to learn from few examples. *arXiv preprint arXiv:1903.03096*, 2019.