

Supplemental Material for Enhancing Few-Shot Image Classification with Unlabelled Examples

Peyman Bateni^{1,6‡§}, Jarred Barber^{2*¶}, Jan-Willem van de Meent³, Frank Wood^{1,4,5}
University of British Columbia¹, Amazon², Northeastern University³, Inverted AI⁴, MILA⁵, Beam AI⁶
pbateni@cs.ubc.ca, jarred.barber@gmail.com, j.vandemeent@northeastern.edu, fwood@cs.ubc.ca

Appendices

A. Benchmarks and Training

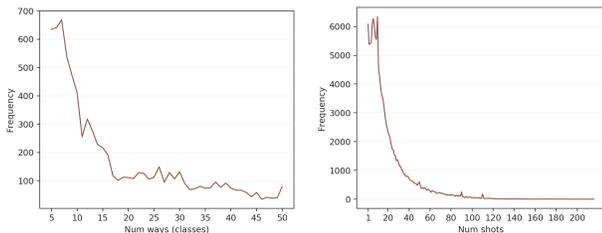
A.1. Meta-Dataset

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 [53], and we refer the interested reader to Section 3.2 in [53] 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 7a, and the class frequency as compared to the class shot is presented in Figure 7b. 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

‡ Authors contributed equally

§ Work performed while at Inverted AI

¶ Work performed while at Charles River Analytics



(a) Number of Tasks vs. Ways

(b) Number of Classes vs. Shots

Figure 7: Test-time way and shot frequency graphs. Figure is directly from Bateni et al. [2]. As shown, most tasks have fewer than 10 classes (way) and most classes have less than 20 support examples (shot).

total images to support 10 query examples.

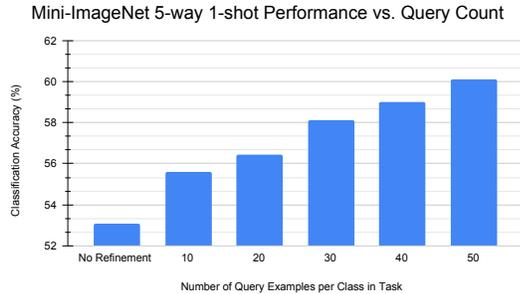
A.2. mini/tiered-ImageNet

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.

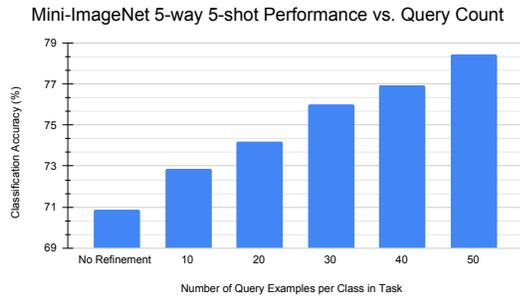
A.3. Meta-Dataset Training/Testing

Following [2] and [41], we train our ResNet18 feature extractor as a supervised multi-class image classifier on the training split of the ImageNet subset of the Meta-Dataset. We directly follow the procedure described in C.1.1 of [41] and use their released checkpoint. Images from the 712 ImageNet classes designated for training by Meta-Dataset [53] 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 [48, 7] 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



(a) 5-way 1-shot



(b) 5-way 5-shot

Figure 8: Mini-ImageNet performance as the number of query examples per class increases on 1/5-shot 5-way tasks.

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 [42] 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.

B. Additional Experiments

B.1. Performance vs. Number of Query Examples

Figure 8 shows the performance of Transductive CNAPS on 1/5-shot 5-way Mini-ImageNet tasks as the number of

query examples per category increases from 10 to 50. As shown, with a greater number of query examples, performance improves as our method is able to exploit more unlabelled data.