

A. Training Details

To train and test MetaNet on one-shot learning, we adapted the training procedure introduced by Vinyals et al. (2016). First, we split the data into training and test sets consisting of two disjoint classes. We then formulate a series of tasks (trials) from the training set. Each task has a support set of N classes with one image per, resulting in an N -way one-shot classification problem. In addition to the support set, we also include L number of labeled examples in each task set to update the parameters θ during training. For testing, we follow the same procedure to form a set of test tasks from the disjoint classes. However, now MetaNet assigns class labels to L examples based only on the labeled support set of each test task.

For the one-shot benchmarks on the Omniglot dataset, we used a CNN with 64 filters as the base learner b . This CNN has 5 convolutional layers, each of which is a 3×3 convolution with 64 filters, followed by a *ReLU* non-linearity, a 2×2 max-pooling layer, a fully connected (FC) layer, and a softmax layer. Another CNN with the same architecture is used to define the dynamic representation learning function u , from which we take the output of the FC layer as the task dependent representation r . We trained a similar CNNs architecture with 32 filters for the experiment on Mini-ImageNet. However for computational efficiency as well as to demonstrate the flexibility of MetaNet, the last three layers of these CNN models were augmented by fast weights. For the networks d and m , we used a single-layer LSTM with 20 hidden units and a three-layer MLP with 20 hidden units and *ReLU* non-linearity. As in Andrychowicz et al. (2016), the parameters G and Z of d and m are shared across the coordinates of the gradients ∇ and the gradients are normalized using the same preprocessing rule (with $p = 7$). The MetaNet parameters θ are optimized with ADAM. The initial learning rate was set to 10^{-3} . The model parameters θ were randomly initialized from the uniform distribution over $[-0.1, 0.1]$.

B. MNIST as Out-Of-Domain Data

We treated MNIST images as a separate domain data. Particularly a model is trained on the Omniglot training set and evaluated on the MNIST test set in 10-way one-shot learning setup. We hypothesize that models with a high dynamic should perform well on this task.

In Figure 5, we plotted the results of this experiment. MetaNet- achieved 71.6% accuracy which was 0.6% and 3.2% lower than the other variants with fast weights. This is not surprising since MetaNet without dynamic representation learning function lacks an ability to adapt its parameters to MNIST image representations. The standard MetaNet model achieved 74.8% and MetaNet+ obtained

72.3%. Matching Net (Vinyals et al., 2016) reported 72.0% accuracy in this setup. Again we did not observe improvement with MetaNet+ model here. The best result was recently reported by using a generative model, Neural Statistician, that extends variational autoencoder to summarize input set (Edwards & Storkey, 2017).

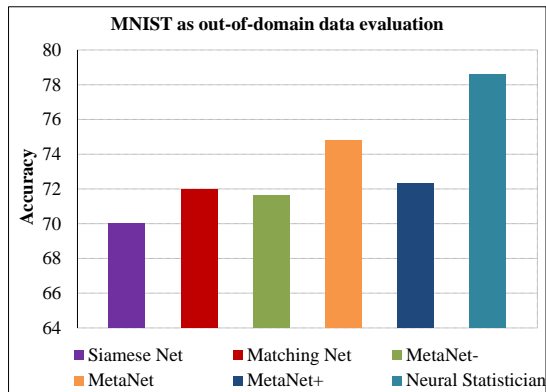


Figure 5. MNIST 10-way one-shot classification results.