Meta-Learner with Linear Nulling: Supplementary Material

A Details of Learning and Classification Procedures

Algorithm 1 provides detailed steps of the initial learning procedure of our meta-learner. For each training episode, \( N_c \) classes are randomly chosen from the training set of a given dataset. Then, for each class, \( N_s \) labeled samples are randomly chosen as the support set \( S_k \), and \( N_q \) labeled samples are chosen as the query set \( Q_k \), without any overlapping samples between \( S_k \) and \( Q_k \). With the support set \( S_k \), the average network output vector \( \bar{g}_k \) is obtained for each class (in line 5). Based on the per-class average network output vectors, error vectors are obtained for all classes (in line 6) without any relabeling on the reference vectors. Then the linear transformer \( M \) is computed as a null-space of the error signals. For each query input, the Euclidean distances to the reference vectors in the projection space \( M \) are measured, and the training loss is computed using these distances. The average training loss is obtained over all \( N_q \) query inputs of \( N_c \) classes (in line 11 to 14). The learnable parameters \( \theta \) of the embedding network and the references \( \Phi \) are now updated with the average training loss (in line 16).

B Hyperparameters in Experiment

In Table 1 we show the hyperparameters used for 20-way Omniglot and 5-way miniImageNet experiments in the main paper. For all experiments, the initial learning rate is \( 10^{-3} \), but the rate decays by half in every \( S_d \) episodes in the miniImageNet experiments. \( S_d \), the learning rate decay step, and \( N_q \), the number of query images per class in each episode, are chosen empirically.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>( S_d )</th>
<th>( N_q )</th>
</tr>
</thead>
<tbody>
<tr>
<td>20-way Omniglot 1-shot</td>
<td>No decay</td>
<td>7</td>
</tr>
<tr>
<td>20-way Omniglot 5-shot</td>
<td>No decay</td>
<td>7</td>
</tr>
<tr>
<td>5-way miniImageNet 1-shot</td>
<td>5000</td>
<td>5</td>
</tr>
<tr>
<td>5-way miniImageNet 5-shot</td>
<td>7500</td>
<td>2</td>
</tr>
</tbody>
</table>
Algorithm 1: Initial learning is done by $N_E$ training episodes. Each episode $E_i$ consists of $N$ (image, label) pairs. These $N$ shots are composed of $N_c$ classes and there are $N_s$ shots and $N_q$ queries in each class. $L_{train}$ is the loss for training learnable parameters. The Euclidean distance between two vectors is denoted as $d(\cdot, \cdot)$.

Input: Training set $E^T = \{E_1, \ldots, E_{N_E}\}$ where $E_i = \{(x_1, y_1), \ldots, (x_N, y_N)\}$ is an episode with $N = N_c(N_s + N_q)$ pairs of image and label where $y_n \in \{0, \ldots, N_c - 1\}$. $E_i^{(k)} = \{(x_1^{(k)}, y_1^{(k)}), \ldots, (x_{N_s+N_q}^{(k)}, y_{N_s+N_q}^{(k)})\}$ is the subset of $E_i$ consisting of all pairs $(x_n, y_n)$ such that $y_n = k$.

1: for $i$ in $\{1, \ldots, N_E\}$ do
2: \hspace{1em} $L_{train} \leftarrow 0$
3: \hspace{1em} for $k$ in $\{0, \ldots, N_c - 1\}$ do
4: \hspace{2em} $S_k \leftarrow \{(x_n^{(k)}, y_n^{(k)})\}$ with $(x_n^{(k)}, y_n^{(k)}) \in E_i^{(k)}, n \leq N_s$
5: \hspace{2em} $\bar{g}_k \leftarrow \frac{1}{N_s} \sum_{(x_n^{(k)}, y_n^{(k)}) \in S_k} f_\theta(x_n)$
6: \hspace{2em} $v_k \leftarrow \{(N_c - 1)\phi_k - \sum_{l \neq k} \phi_l\} - \bar{g}_k$
7: \hspace{1em} end for
8: \hspace{1em} $M \leftarrow \text{null} \left(\{v_k\}_{k \in \{0, \ldots, N_c - 1\}}\right)$
9: \hspace{1em} for $k$ in $\{0, \ldots, N_c - 1\}$ do
10: \hspace{2em} $Q_k \leftarrow \{(x_n^{(k)}, y_n^{(k)})\}$ with $(x_n^{(k)}, y_n^{(k)}) \in E_i^{(k)}, N_s < n \leq N_s + N_q$
11: \hspace{2em} for $(x_q, y_q)$ in $Q_k$ do
12: \hspace{3em} $g_q \leftarrow f_\theta(x_q)$
13: \hspace{2em} $L_{train} \leftarrow L_{train} + \frac{1}{N_c N_q} \left[ d(\phi_k M, g_q M) + \log \sum_{k'} \exp(-d(\phi_{k'} M, g_q M)) \right]$ + $\log \sum_{k'} \exp(-d(\phi_{k'} M, g_q M))$
14: \hspace{1em} end for
15: \hspace{1em} end for
16: Update $\theta, \Phi$ minimizing $L_{train}$ via Adam optimizer
17: end for