Strategies for Meta-Learning with Diverse Tasks

Stefano Woerner STEFANO.WOERNER@UNI-TUEBINGEN.DE
Christian F. Baumgartner CHRISTIAN.BAUMGARTNER@UNI-TUEBINGEN.DE
Cluster of Excellence Machine Learning: New Perspectives for Science, University of Tübingen

Editors: Under Review for MIDL 2022

Abstract

A major limitation of deep learning for medical applications is the scarcity of labelled data. Meta-learning, which leverages principles learned from previous tasks for new tasks, has the potential to mitigate this data scarcity. However, most meta-learning methods assume idealised settings with homogeneous task definitions. The most widely used family of meta-learning methods, those based on Model-Agnostic Meta-Learning (MAML), require a constant network architecture and therefore a fixed number of classes per classification task. Here, we extend MAML to more realistic settings in which the number of classes can vary by adding a new classification layer for each new task. Specifically, we investigate various initialisation strategies for these new layers. We identify a number of such strategies that substantially outperform the naive default (Kaiming) initialisation scheme.

1. Introduction

While deep learning approaches have led to breakthroughs in many medical image analysis tasks they are notoriously hindered by the availability of large data sets. The collection of such data sets is often difficult and costly, and in some cases altogether prohibitive such as in the case of very rare diseases.

Meta-learning is a class of techniques which tries to mimic the human ability to learn new tasks very quickly by leveraging life-long experience with learning diverse tasks. Specifically, meta learning techniques first “learn to learn” from a set of training tasks before learning the desired test task. Indeed, it has been shown that meta-learning approaches are able to learn previously unseen tasks with very few examples (Finn et al., 2017; Makarevich et al., 2021). These methods hold the potential to bring the predictive power of deep networks to medical domains where they were previously impractical to use due to limited data. However, the vast majority of meta-learning methods assume idealised settings with homogeneous task definitions hindering applications to realistic medical tasks. Specifically, most prior work assumes that all tasks have a constant number of K classes (Finn et al., 2017; Ye and Chao, 2021). In reality, of course, tasks typically have a varying number of target labels K. For the most commonly used family of methods, i.e. methods based on Model-Agnostic Meta-Learning (MAML) (Finn et al., 2017), the restriction that the number of tasks needs to be constant derives directly from the requirement of a constant network architecture. While some works have shown the feasibility of MAML-based methods in medical settings (e.g. (Makarevich et al., 2021)), the above limitations have thus far not been addressed.

In this work, we take a first step in the direction of making meta-learning, and in particular MAML, suitable for more realistic medical problems by investigating different strategies for training and testing with variable K's. Due to the absence of suitable medical benchmarks, we evaluate the method on two commonly used machine learning data sets.

2. Methods

The base method for our experiments is MAML (Finn et al., 2017). The training of MAML involves an inner and an outer training loop. In the outer loop, a batch of tasks is sampled from the meta data set. Those tasks are then trained for a fixed number of steps (inner loop optimisation). The loss for the outer loop optimisation is then provided by the validation score of the trained model on the sampled task obtained in the inner loop.

MAML assumes a constant network architecture for all inner and outer optimisations. In this work, we propose attaching a new classification layer for each iteration such that the number of outputs matches the number of classes for the respective task. Our main contribution lies in exploring strategies for initialising these layers, which are as follows.

Naive Initialisations: Firstly, we investigate re-initialising the weights of the classification layer with the default (i.e. Kaiming) initialisation scheme. Secondly, we investigate initialising the classification layer with zeroes.

Adaptive Initialisation: Next, we introduce a novel initialisation scheme in which we sample the new classifier's weights from the previous classifiers weight distribution. Specifically, at the end of every outer loop step, we save $\mu_{\mathbf{W}} := \mu[(W_{k,j})_{k,j}], \sigma_{\mathbf{W}} := \sigma[(W_{k,j})_{k,j}]$, as well as $\mu_{\boldsymbol{b}} := \mu[(b_k)_k], \sigma_{\boldsymbol{b}} := \sigma[(b_k)_k]$ and initialise with $(W_{k,j})_{k,j} \sim \mathcal{N}(\mu_{\mathbf{W}}, \sigma_{\mathbf{W}}), (b_k)_k \sim \mathcal{N}(\mu_{\boldsymbol{b}}, \sigma_{\boldsymbol{b}})$ at the beginning of the next episode.

Unicorn-MAML: Ye and Chao (2021) proposed a variation of MAML which is invariant to different permutations of the class labels, which they coined Unicorn-MAML. The approach relies on learning a single weight vector $\boldsymbol{w}$ and bias b in the outer loop instead of separate weights per class. At the beginning of each episode the parameters are set to $\boldsymbol{W}_k = \boldsymbol{w}$ and $b_k = b$. While the approach was not intended to work with variable K's it can be trivially extended to this end by repeating $\boldsymbol{w}$ the required number of times.

Proto-MAML: Triantafillou et al. (2020) introduced Proto-MAML which combines MAML with prototypical networks. Prototypical networks classify samples according to the minimum distance to prototypes (class centroids) derived from the support set. With prototypes $(\boldsymbol{c}_k)_1^K$ the weights and biases of the classification layer can be set to $\boldsymbol{W}_k = 2\boldsymbol{c}_k$ and $b_k = -||\boldsymbol{c}_k||^2$. Again, this approach can be trivially extended for a varying number of outputs as the required centroids can be generated for any support set on the fly.

3. Experiments & Results

We evaluate all approaches on the *MiniImageNet* and *TieredImageNet* data sets which are widely used in the meta learning community (see e.g. (Finn et al., 2017; Ye and Chao, 2021)). For all experiments, we train with $K = 5$ and 5 labelled examples. However, for meta test we randomly uniformly sample K from $\{3, 4, 5, 6, 7\}$. The training for all methods was performed as described in (Finn et al., 2017). The results are shown in Tab. 1.

One of our main observations is that the perhaps most "obvious" strategy, i.e. reinitialising the new classifier with the default initialisation method, leads to a considerably lower performance compared to the other reference methods. We furthermore conclude that all other initialisation schemes appear to perform more or less en-par with each other. Perhaps surprisingly, a simple initialisation with zeroes performs better than (MiniImageNet) or en par with (TieredImageNet) more sophisticated schemes. While Proto-MAML has a slight

Table 1: Meta-test accuracy in percent (with 0.95 confidence interval).

Method	*MiniImageNet*	*TieredImageNet*
Default Initialisation (Kaiming)	59.68 ± 1.21	64.67 ± 1.17
Initialization with zeroes	$\mathbf{65.81 \pm 1.14}$	69.29 ± 1.16
Adaptive Initialisation	64.36 ± 1.14	69.55 ± 1.13
Unicorn-MAML	65.38 ± 0.97	69.79 ± 1.03
Proto-MAML	64.80 ± 0.98	$\mathbf{70.35 \pm 1.09}$

advantage on the more challenging TieredImageNet data set, the computation overhead of obtaining the class centroids may not justify the marginal benefit over simpler techniques. Looking at the adaptive sampling scheme in detail we noticed that the mean and standard deviation both converge to zero, essentially leading to an initialisation with permutation invariance. As a consequence all strategies, except the default initialisation, have the property of permutation invariance. We thus hypothesise that permutation invariance is a crucial ingredient for a variable K initialisation scheme. As a practical recommendation we suggest initialising the classification layer of MAML-based methods with zeroes for applications with variable K, due to the easy implementation and lack of computational overheads.

Due to the lack of suitable meta-learning benchmarks for medical image analysis we had to fall back on standard benchmarks from the meta-learning literature. In future work we plan to create such a benchmark and confirm our findings.

Acknowledgments

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC number 2064/1 – Project number 390727645. The authors thank the International Max Planck Research School for Intelligent Systems (IMPRS-IS) for supporting Stefano Woerner.

References

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, pages 1126–1135. PMLR, 06–11 Aug 2017.

Anastasia Makarevich, Azade Farshad, Vasileios Belagiannis, and Nassir Navab. MetaMedSeg: Volumetric meta-learning for few-shot organ segmentation. *arXiv preprint arXiv:2109.09734*, 2021.

Eleni Triantafillou, Tyler Zhu, Vincent Dumoulin, Pascal Lamblin, Utku Evci, 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. In *International Conference on Learning Representations*, 2020.

Han-Jia Ye and Wei-Lun Chao. How to train your MAML to excel in few-shot classification. *arXiv preprint arXiv:2106.16245*, June 2021.