

Supplementary Material for Adaptive Task Sampling and Variance Reduction for Gradient-Based Meta-Learning

Zhuoqun Liu
lzqyoooh@sjtu.edu.cn

Yuankun Jiang
yuankunjiang@sjtu.edu.cn

Chenglin Li
lcl1985@sjtu.edu.cn

Wenrui Dai
daiwenrui@sjtu.edu.cn

Junni Zou
zoujunni@sjtu.edu.cn

Hongkai Xiong
xionghongkai@sjtu.edu.cn

Shanghai Jiao Tong University
800 Dongchuan Road,
Shanghai 200240, China

1 Evaluation on Computational Complexity

In this supplementary material, we compare the computational complexity of our proposed ATSVR with the other two state-of-the-art meta-learning methods ANIL [1] and ATS [2], in terms of the model size, training time, and inference time. Note that ANIL [1] is compared here, because it is the backbone meta-learner of both ATS and ATSVR. While ATS [2] is chosen for comparison since it achieves the second best few-shot image classification performance as shown in Tables 2-4 in the main text. Here, we test all different models on a machine with an NVIDIA Tesla V100 graphics card, and with the CPU thread count set to 4. We follow the same settings of network architecture as in Section 5.2 of the main text, and show in Table 1 the comparison results on the miniImageNet dataset with limited budget 16, 5-way 5-shot setting. Note that for the model size, ATSVR requires additionally a pretrained recurrent aggregator $\mathcal{A}(\mathcal{F}(\cdot))$, which may incur additional parameters depending on its network architecture (about 180K in our experiment), but will not notably affect the training and inference time. For the training time, we compare the time required to train one iteration when the meta batchsize is equal to 1, while we test the inference time of three models on one single task.

It can be seen that with the same 4-layer 32-channel CNN setting for the meta learner, the proposed network to learn the target distribution $q(\mathcal{T}; \phi)$ only introduces a slightly larger model parameters than ANIL [1], also with relatively longer training. This can be expected

since ANIL is designed for reducing the training time. However, compared to ATS [10] which achieves the second-best performance in the few-shot classification, our training time reduces significantly, which is mainly because instead of adjusting the sampling distribution w.r.t. the task difficulty, we can adjust it directly in the task space.

	Number of Parameters	Training Time	Inference Time
ANIL [10]	32.77K	64.18 ms	19.49 ms
ATS [10]	40.07K	915.71 ms	19.77 ms
ATSVR	39.61K	80.64 ms	19.73 ms

Table 1: Comparison on the number of parameters, training time per task, and inference time per task achieved by different methods, where all the meta learners set their networks to a 4-layer 32-channel CNN.

References

- [1] Aniruddh Raghu, Maithra Raghu, Samy Bengio, and Oriol Vinyals. Rapid learning or feature reuse? towards understanding the effectiveness of maml. *arXiv preprint arXiv:1909.09157*, 2019.
- [2] Huaxiu Yao, Yu Wang, Ying Wei, Peilin Zhao, Mehrdad Mahdavi, Defu Lian, and Chelsea Finn. Meta-learning with an adaptive task scheduler. *CoRR*, abs/2110.14057, 2021. URL <https://arxiv.org/abs/2110.14057>.