Towards Robust Episodic Meta-Learning: Supplementary Material

Beyza Ermis[1] **Giovanni Zappella**[1] **Cédric Archambeau**[1]

[1]Amazon Web Services, Berlin, Germany

A APPENDIX

A.1 SHOTS ANALYSIS

The results of active query and/or context set selections for varying episode sizes $T = \{40, 100, 200, 400\}$ when $N = 20$ is presented in figures below.

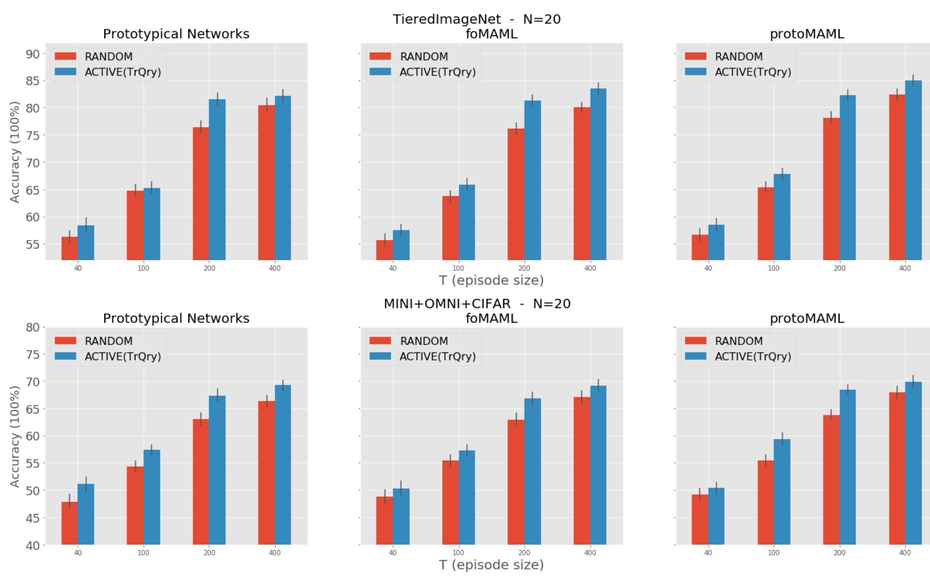

Figure 1: Comparison of random selection (RANDOM) and active selection (ACTIVE) for query set construction during meta-train (TrQry) for varying episode sizes T when the way $N = 20$ on TieredImageNet and MINI+OMNI+CIFAR.

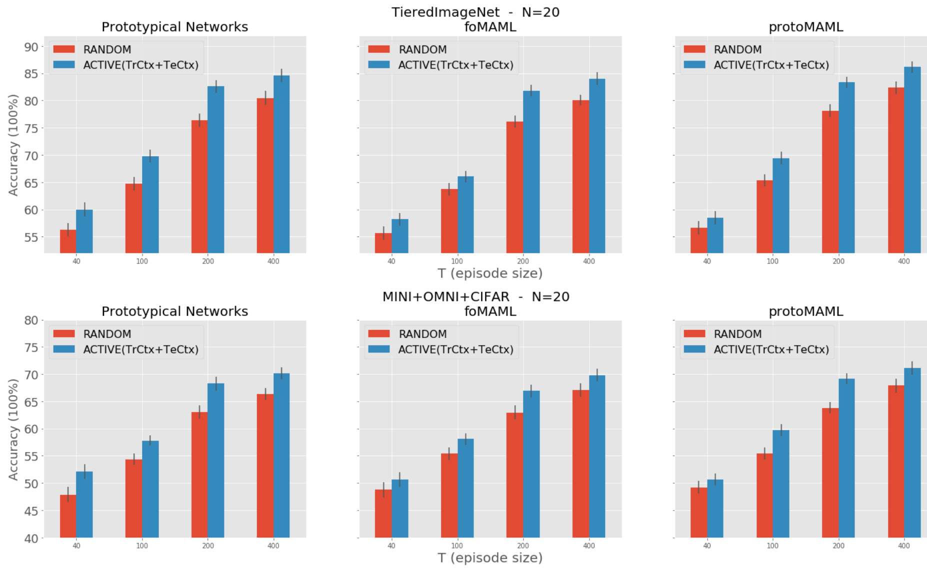

Figure 2: Comparison of random selection (RANDOM) and active selection (ACTIVE) during meta-train and meta-test for the generation of the context set (TrCtx+TeCtx) for varying episode sizes T when the way $N = 20$ on TieredImageNet and MINI+OMNI+CIFAR.

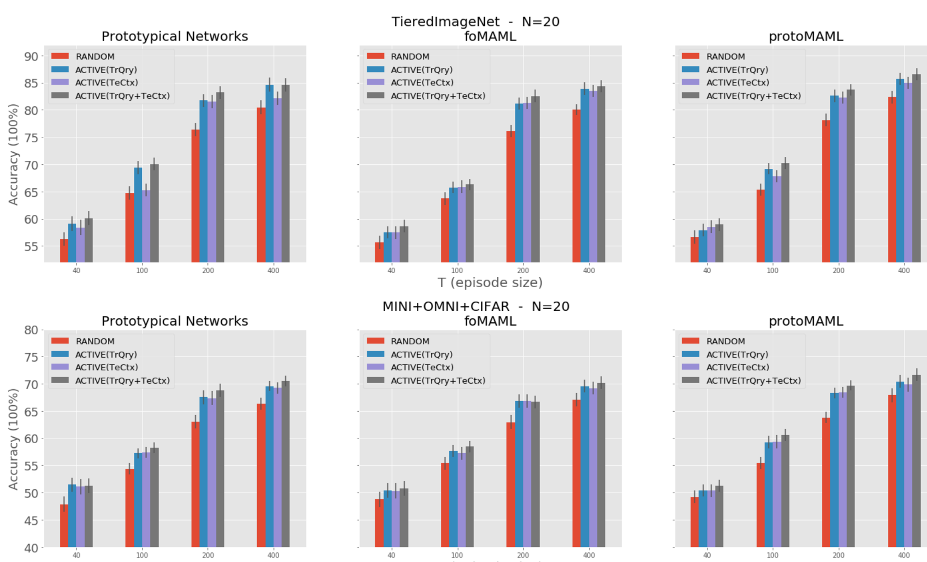

Figure 3: Comparison of random selection (RANDOM) with active selection (called ACTIVE) for the creation of context set at meta-test time (TeCtx), query set at meta-train time (TrQry), and their combination for varying episode sizes T when the way $N = 20$ on TieredImageNet and MINI+OMNI+CIFAR.

A.2 CIFAR100 RESULTS

In this set of experiments, only tasks from CIFAR100 are used both in train and test.

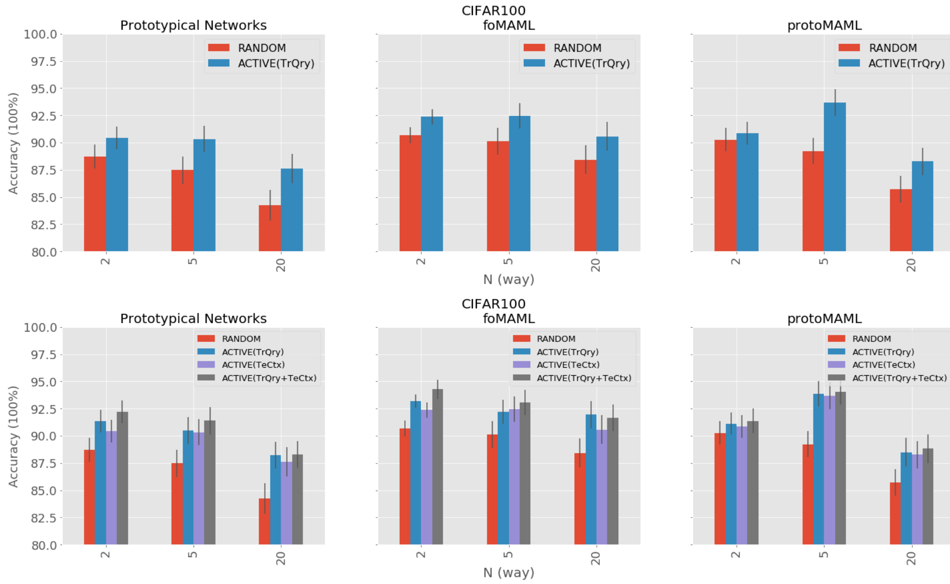

Figure 4: Comparison of random selection (RANDOM) and active selection (ACTIVE) for query set construction, context set construction and their combination for varying N when the episode sizes T are $N \times 10$ on CIFAR100.

As expected, a bigger improvement is visible in the more complex experiments (that we presented in experiments section), TieredImageNet and MINI+OMNI+CIFAR where all datasets are mixed, but a visible advantage is present also in the simpler scenario where only CIFAR100 is considered.

A.3 ADAPTATION WITH PRE-TRAINED MODELS

An interesting questions arising from our experiments is: can we just use a pre-trained model and improve its performance with by using meta-learning? A possibility already shown in the main text is to employ active learning during at meta-test time for the creation of better context sets. In this section we report a number of experiment showing the advantages provided by this technique.

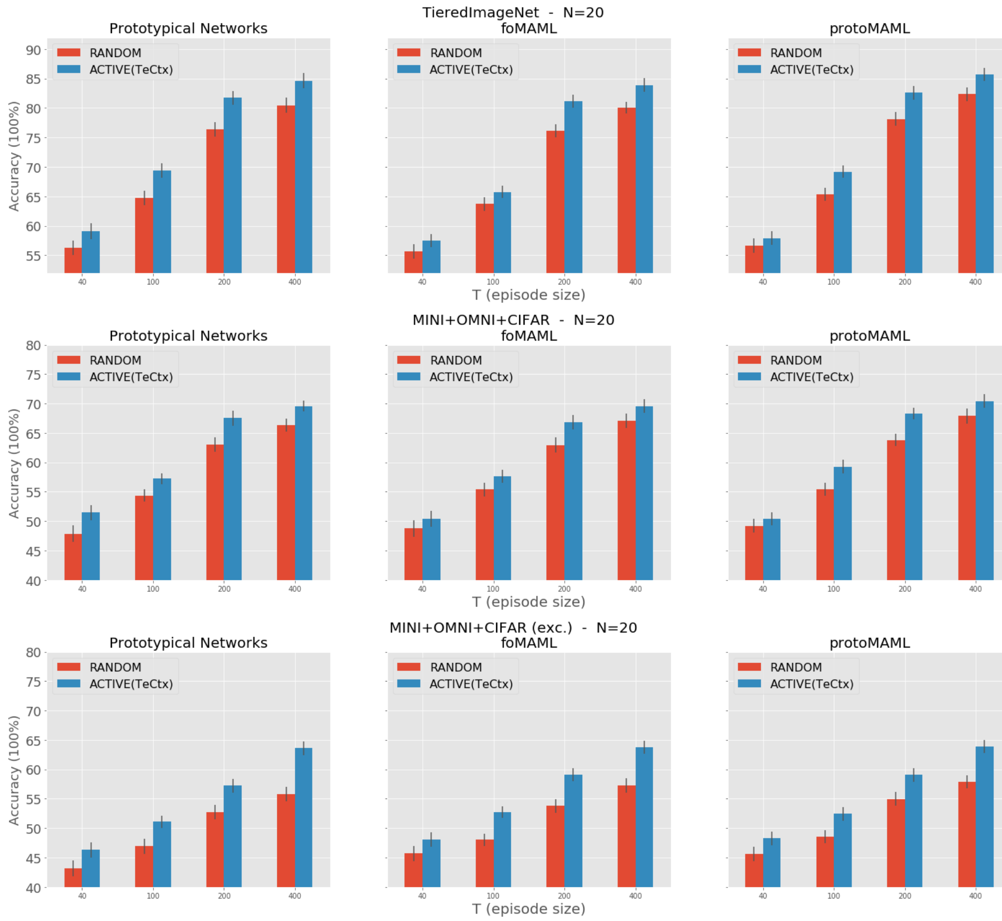

Figure 5: Comparison of random selection and active selection for context set at only meta-test (TeCtx) for varying episode sizes T when the way $N = 20$ on TieredImageNet, MINI+OMNI+CIFAR and MINI+OMNI+CIFAR (excluded).

Our results show that leveraging active learning with pre-trained meta-learning model is beneficial. One interesting finding in that setting is, using the actively selected context sets performs slightly better than the case where we use all the samples in the context sets. Active selection during meta-test also improves the performance of foMAML compared to using all the samples. Because of task homogeneity assumption limitation, MAML does not perform well on the training tasks sampled from different distributions [Vuorio et al., 2019].

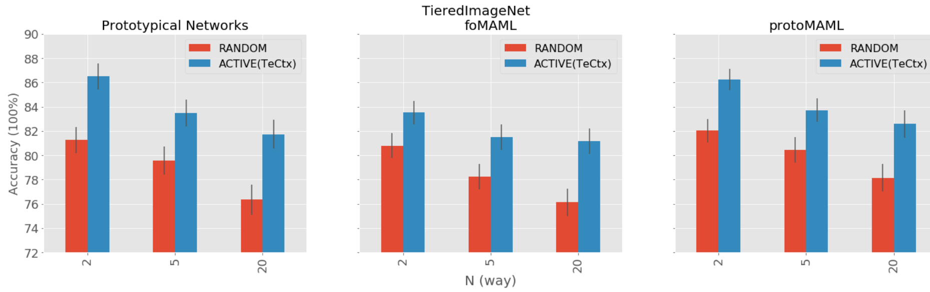

Figure 6: Same comparison as above for the composition of the context set (TrCtx) for varying N when the episode sizes T are $N \times 10$ on TieredImageNet.

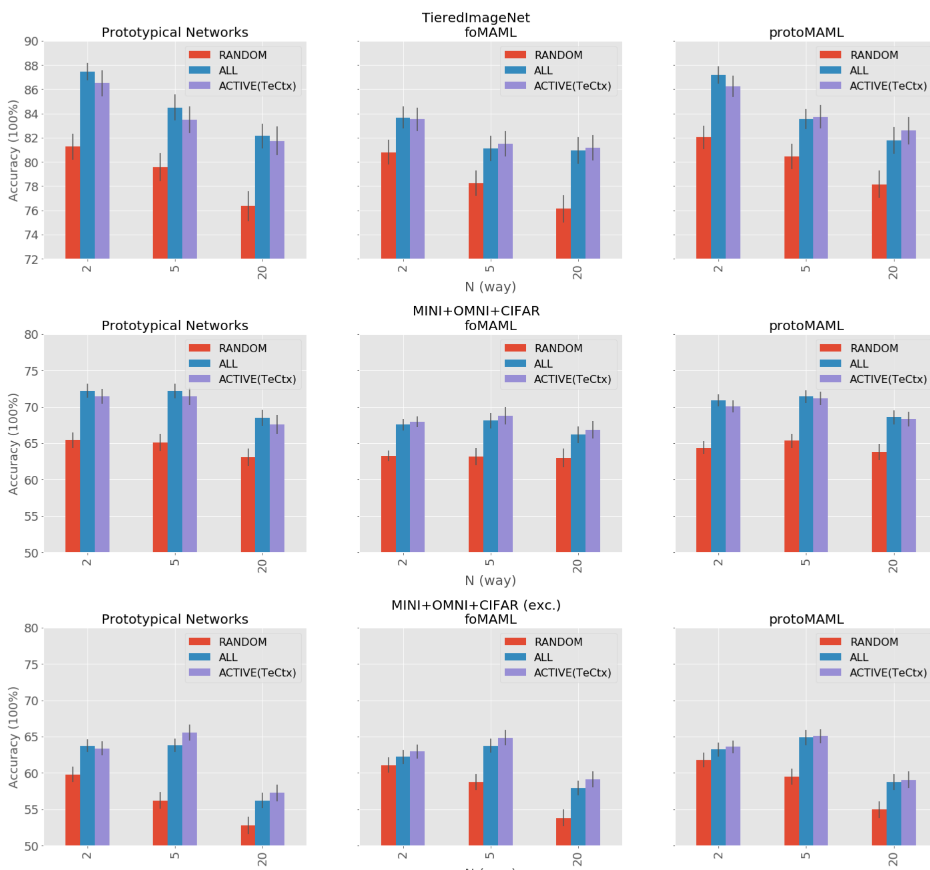

Figure 7: Comparison of random selection (RANDOM), using all the samples in the context set (ALL) and active selection (ACTIVE) during meta-test for the generation of the context set (TrCtx) for varying N when the episode sizes T are $N \times 10$ on TieredImageNet and MINI+OMNI+CIFAR and MINI+OMNI+CIFAR (excluded).

A.4 IS ACTIVE SELECTION WITH UNSTRATIFIED MORE ROBUST?

In this section, we test all the algorithms in a fixed $N = \{2, 5, 20\}$-way and $K = \{2, 10\}$-shot setting (stratified) and compare the results with the ones presented in experiments section for $T = N \times 2$ and $T = N \times 10$ (unstratified). The results show that although the accuracy for both random and active selection strategy is higher in stratified setting; the improvement by active selection is more significant in unstratified setting.

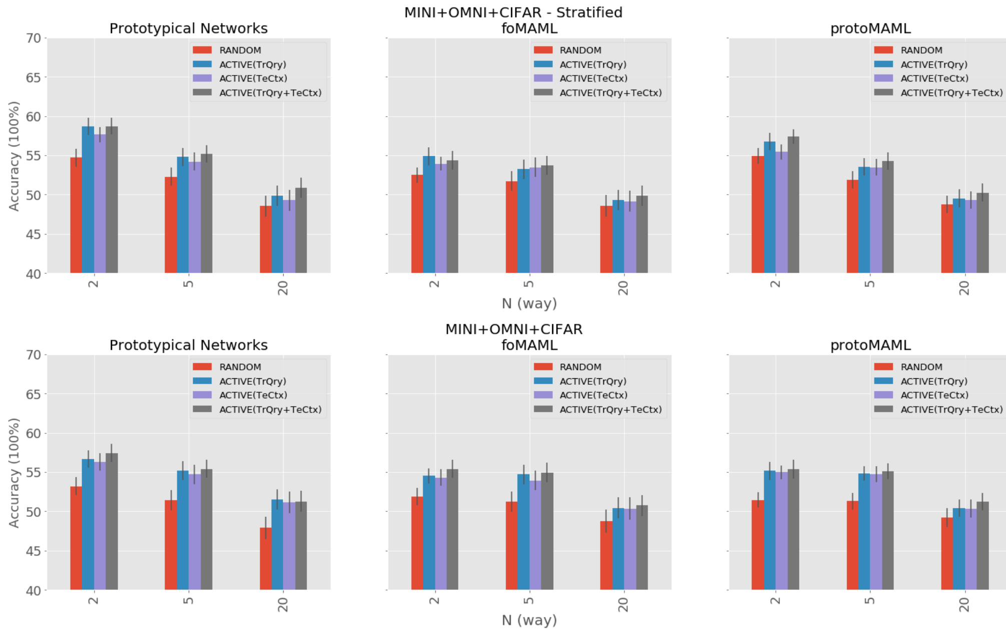

Figure 8: Comparison of random selection (RANDOM) with active selection (called ACTIVE) for the creation of context set at meta-test time (TeCtx), query set at meta-train time (TrQry), and their combination for varying N in stratified setting when $K = 2$ (top figure) and in unstratified setting when $T = N \times 2$ (bottom figure) on MINI+OMNI+CIFAR.

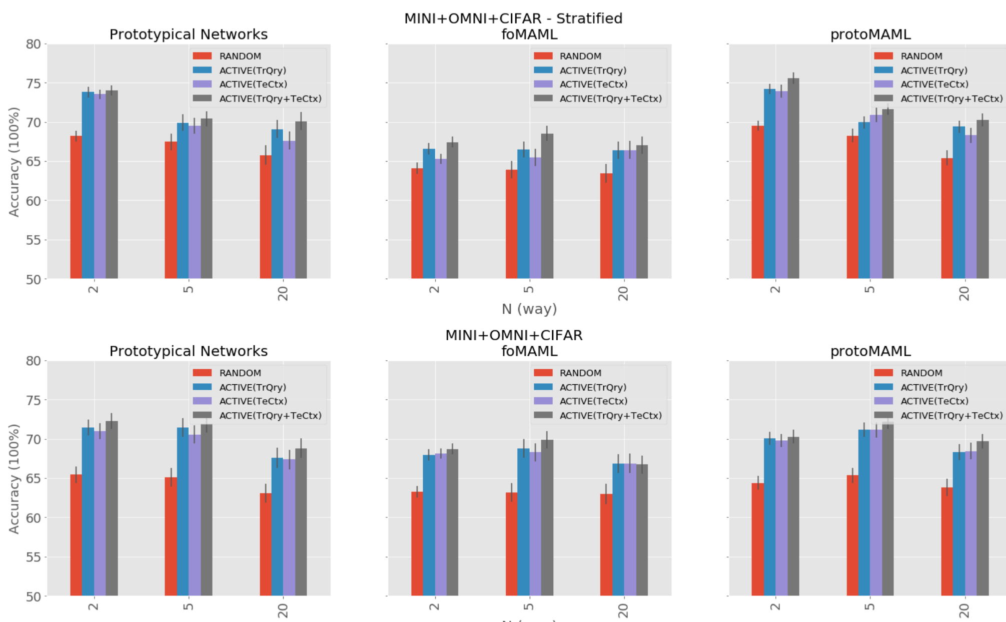

Figure 9: Comparison of random selection (RANDOM) with active selection (called ACTIVE) for the creation of context set at meta-test time (TeCtx), query set at meta-train time (TrQry), and their combination for varying N in stratified setting when $K = 10$ (top figure) and in unstratified setting when $T = N \times 10$ (bottom figure) on MINI+OMNI+CIFAR.

References

Risto Vuorio, Shao-Hua Sun, Hexiang Hu, and Joseph J Lim. Multimodal model-agnostic meta-learning via task-aware modulation. *arXiv preprint arXiv:1910.13616*, 2019.