The Effect of Diversity in Meta-Learning

Ramnath Kumar, Tristan Deleu, Yoshua Bengio

Mila – Quebec Artificial Intelligence Institute

Motivation

- *Few-shot learning*: How to build a model from a small amount of training data?
- Task selection remains virtually unexplored in meta-learning.
- Conventional wisdom says "increasing task diversity would lead to an better performance in models - **We show otherwise!**
- *Applications*: robotics (one-shot imitation learning), healthcare, and other applied research problems.

Problem Statement

- What happens when we change the way tasks are created?
- Large scale analysis of the impact of diversity on the performance in meta-learning:
  - 2 few-shot classification datasets: Omniglot (Lake et al., 2011) & *mini*ImageNet Omniglot (Ravi & Larochelle, 2016)
  - 6 meta-learning algorithms
  - 8 task sampling schemes

Meta-Learning Algorithms

Optimization-based:
- MAML (Finn et al., 2017)
- Reptile (Nichol et al., 2018)

Metric-based:
- Prototypical Networks (Snell et al., 2017)
- Matching Networks (Vinyals et al., 2016)

Bayesian-based:
- CNAPs (Requeima et al., 2019)

Hybrid (Metric + Optimization):
- MetaOptNet (Lee et al., 2019)

Task Sampling Schemes

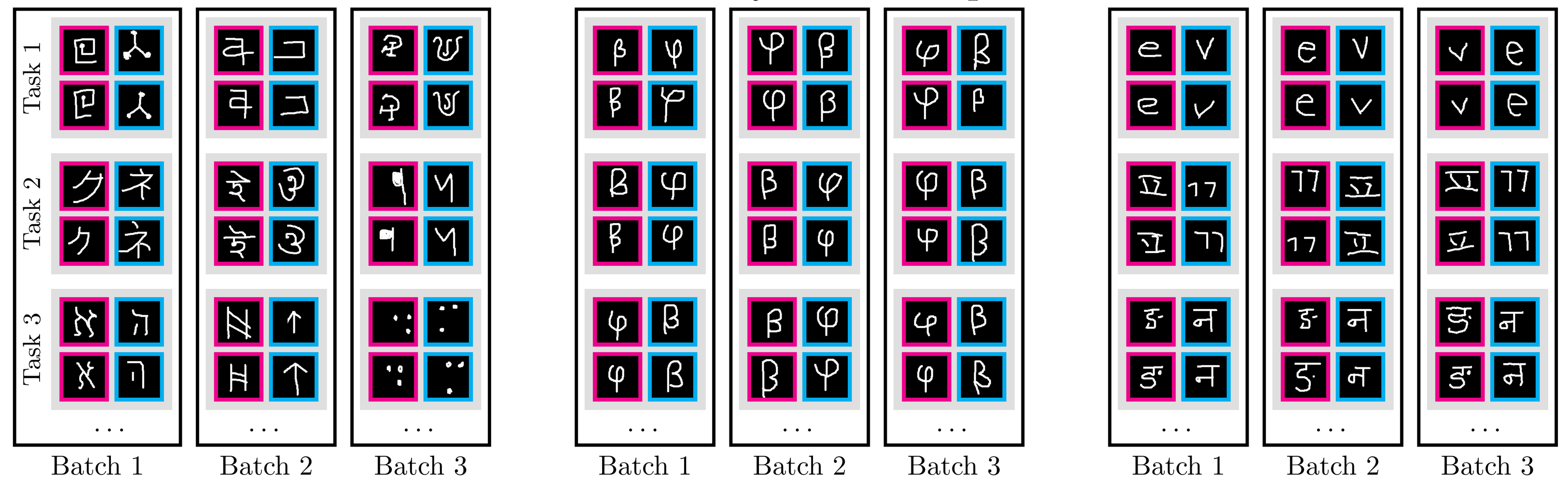
Benchmark - standard sampling scheme:
- Uniform Sampler

Reduced Diversity:
- No Diversity Task Sampler
- No Diversity Batch Sampler
- No Diversity Tasks per Batch Sampler
- Single Batch Uniform Sampler

Increased Diversity:
- Online Hard Task Mining Sampler
- Static DPP Sampler
- Dynamic DPP Sampler

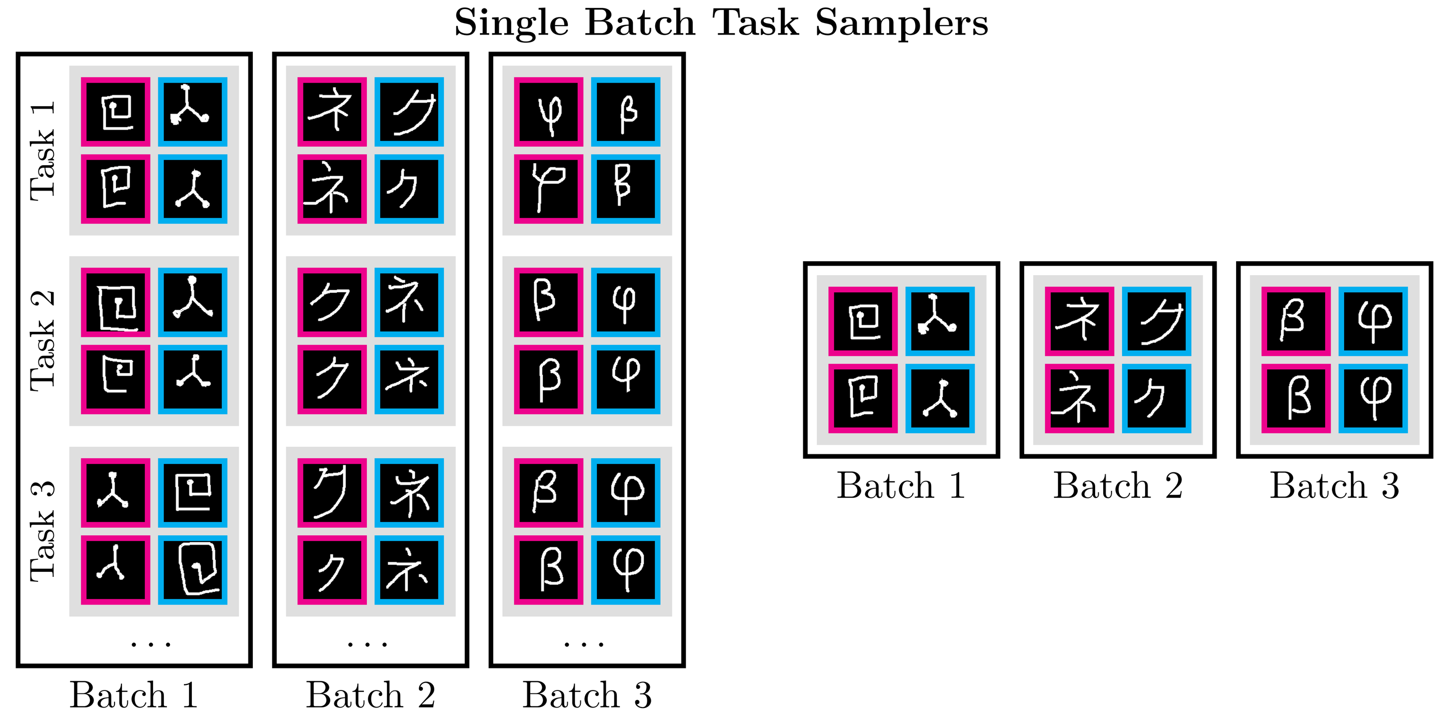
Task Sampling Schemes (contd.)

Low Diversity Task Samplers



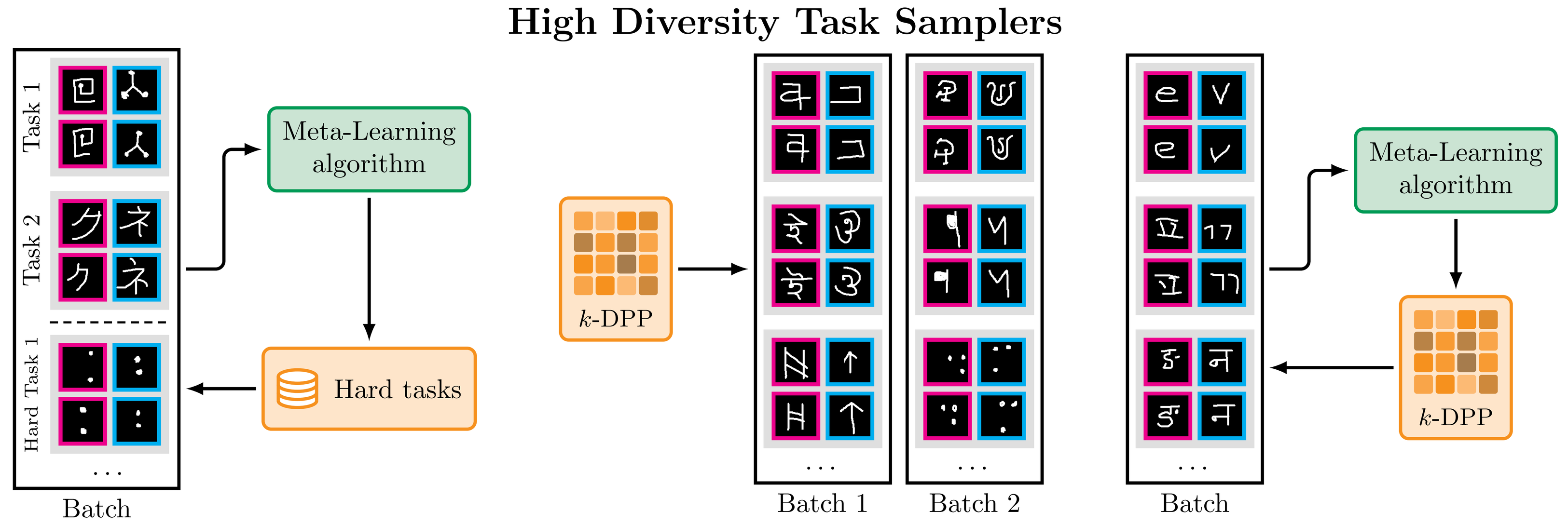
(a) Uniform sampler (b) No Diversity Task sampler (c) No Diversity Batch sampler

Single Batch Task Samplers



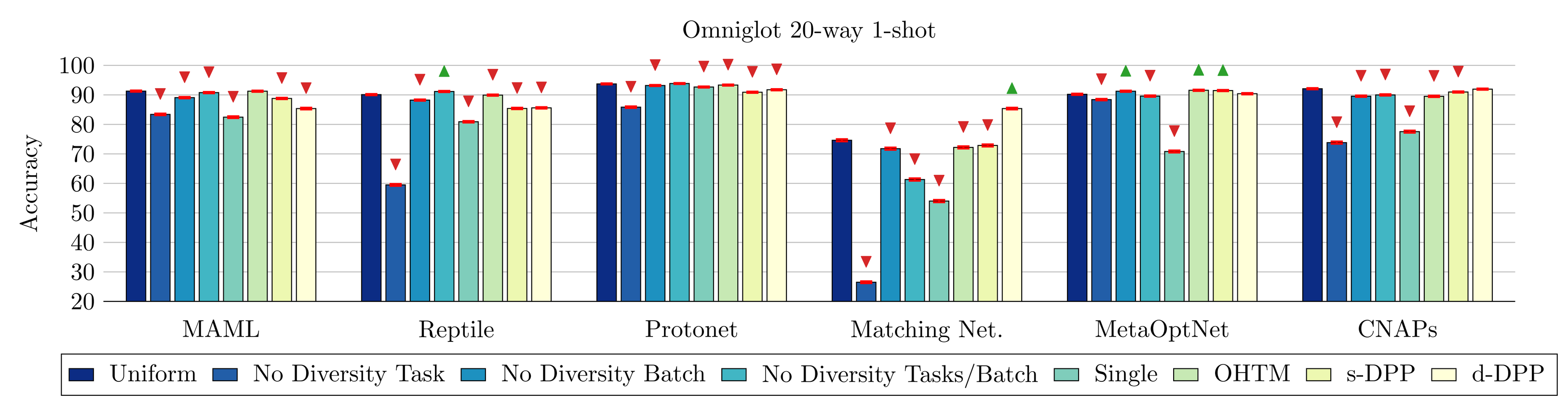
(a) No Diversity Tasks per Batch sampler (b) Single Batch Uniform sampler

High Diversity Task Samplers



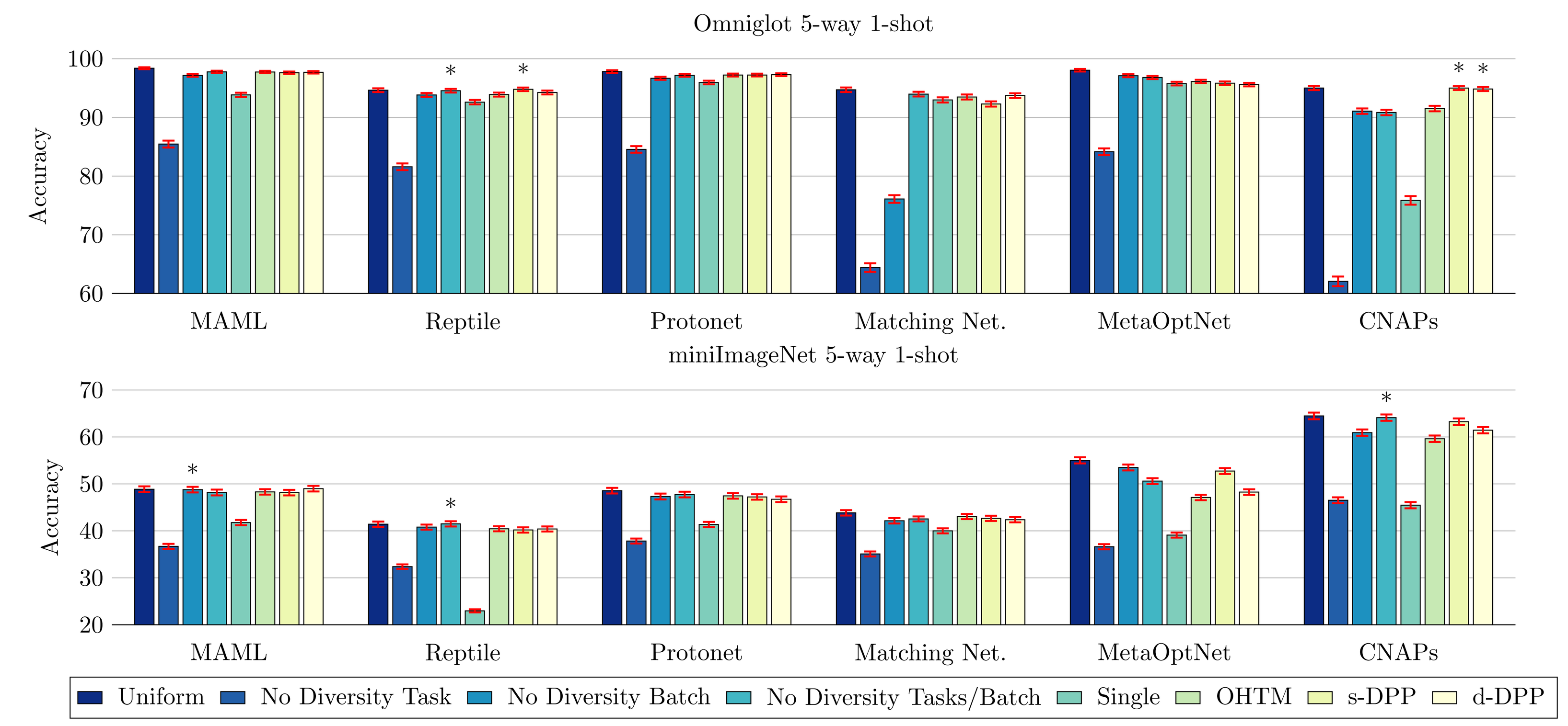
(a) Online Hard Task Mining sampler (b) Static DPP sampler (c) Dynamic DPP sampler

Results



Results on Omniglot 20-way 1-shot

Results (contd.)



Results on Omniglot 5-way 1-shot & *mini*ImageNet 5-way 1-shot

Discussion

- **Poor performance of the NDT Sampler:** (*Expected*), since the model only sees one task. What is fascinating is that just one task is sufficient for the model to reach a reasonably decent performance in most cases.
- **Poor performance of Single vs. NDT/B:** (*Unexpected*), and both should behave the same. However, repeating the same tasks seems to offer some information to the model.
- **OHTM offers no significant performance boost:** (*Unexpected*), since training on hard tasks should improve performance.
- **NDTB, NDB and Uniform Sampler have similar performances:** (*Unexpected*), since we are able to achieve similar results with only marginal amount of data.
- **Disparity between s-DPP and d-DPP:** (*Unexpected*), samplers that improve task diversity usually perform worse and seem to harm the model.

Conclusion

- We identified general trends in this large scale analysis:
  - **High Performing samplers**: No Diversity Batch, No Diversity Task per Batch, Uniform, OHTM, s-DPP.
  - **Low Performing samplers**: No Diversity Task, Single Batch Uniform, d-DPP.
- **Low task-diversity** (eg. No Diversity Batch) does not hurt performance too much - This has impact in practical applications, where uniform sampling is impossible or not efficient.
- **High task-diversity** does not help significantly improve the model, instead it hurts the model.
- This brings into question the efficiency of the model and the advantage it gains with access to more data using samplers such as the standard sampling regime - Uniform Sampler. If we are able to achieve similar performances with less amount of data, the model has not taken advantage of the excess data.

Link to Repository:
`RamnathKumar181/Task-Diversity-meta-learning`