The Effect of Diversity in Meta-Learning

Ramnath Kumar, Tristan Deleu, Yoshua Bengio

Motivation

- Task selection remains virtually unexplored in meta-learning.
- Conventional wisdom is that increasing task diversity would lead to better models.

- The practical implications of our work – 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
  - 8 task sampling schemes
  - 6 meta-learning algorithms

Dataset: Omniglot

- Simple dataset of handwritten characters from various alphabets.
- Contains **1623 different handwritten characters** from **50 different alphabets**
- For each character, we have **20 images**
- Settings: **1-shot 5-way** & **1-shot 20-way**

Mila

Human-level concept learning through probabilistic program induction
Lake, B. M., Salakhutdinov, R., and Tenenbaum, J. B.

Dataset: *mini*ImageNet

- *mini*ImageNet dataset is **extracted from ILSVRC-2012**
- *mini*ImageNet contains **100 classes with 600 images** in each class
- Settings: **1-shot 5-way**

Task Sampling Schemes

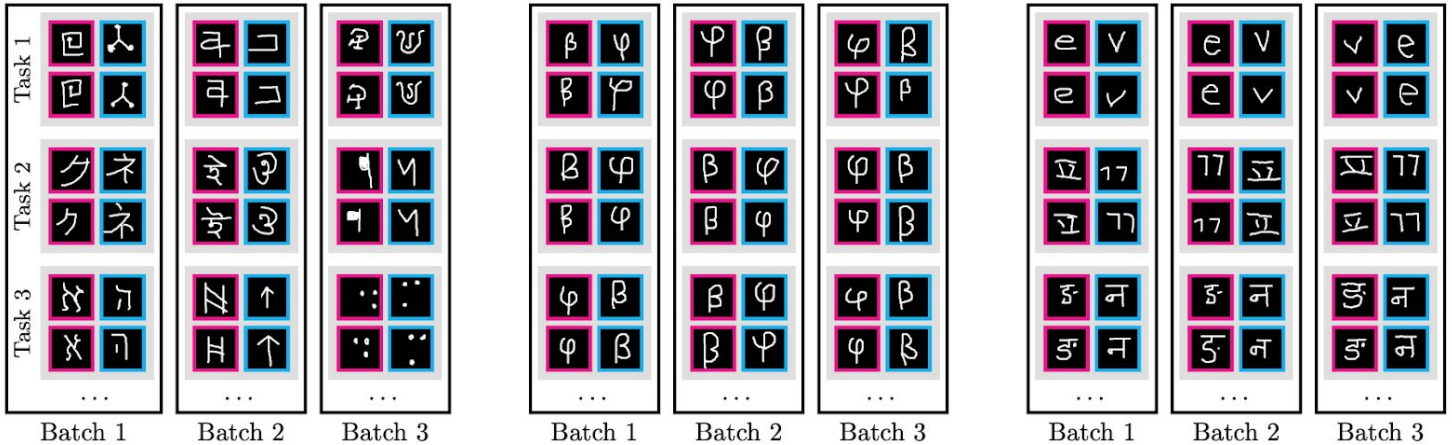
- Uniform Sampler — Our benchmark
- No Diversity Task Sampler
- No Diversity Batch Sampler
- No Diversity Tasks per Batch Sampler
- Single Batch Uniform Sampler

Reduced Diversity

- Online Hard Task Mining Sampler
- Static DPP Sampler
- Dynamic DPP Sampler

Increased Diversity

Task Sampling- Reduced Diversity

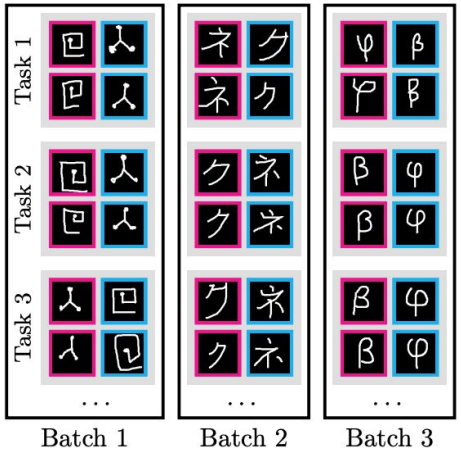


(a) Uniform sampler

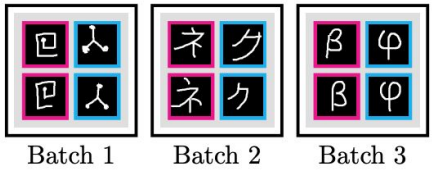
(b) No Diversity Task sampler

(c) No Diversity Batch sampler

Task Sampling- Reduced Diversity

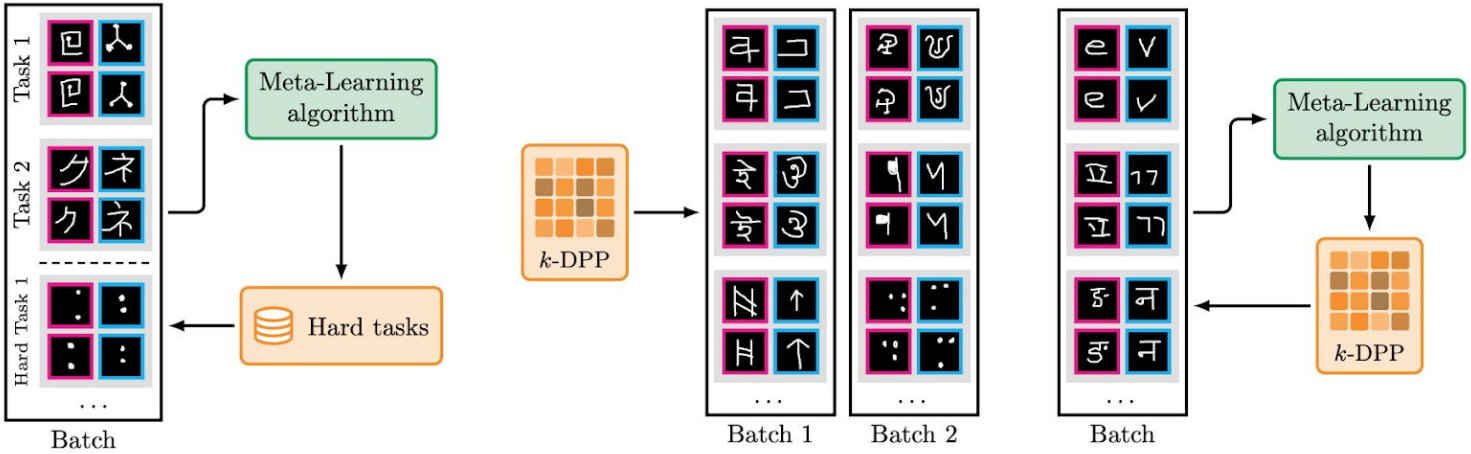


(a) No Diversity Tasks per Batch sampler



(b) Single Batch Uniform sampler

Task Sampling- Increased Diversity



(a) Online Hard Task Mining sampler

(b) Static DPP sampler

(c) Dynamic DPP sampler

Algorithms

- MAML
- Reptile

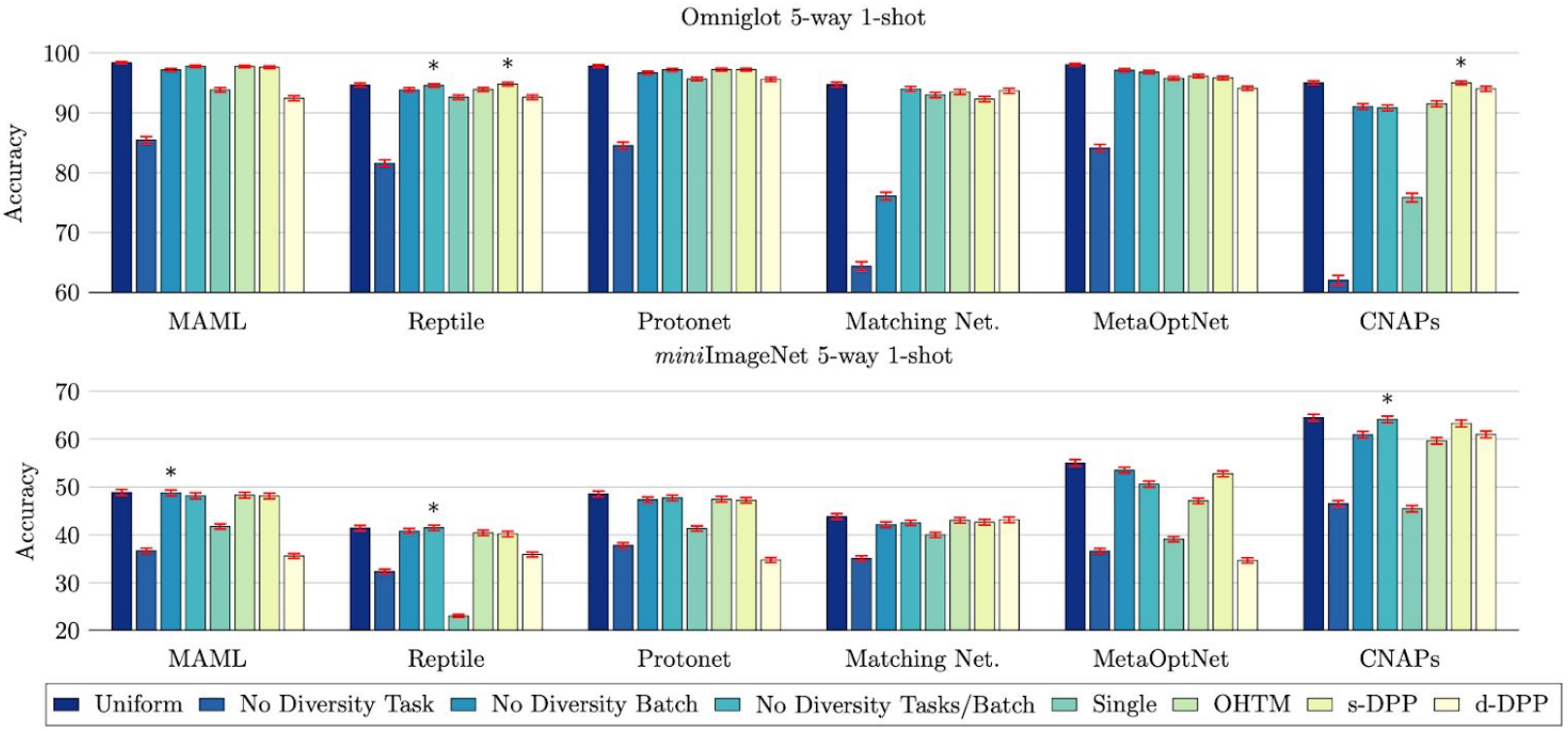
Optimization

- Prototypical Networks
- Matching Networks

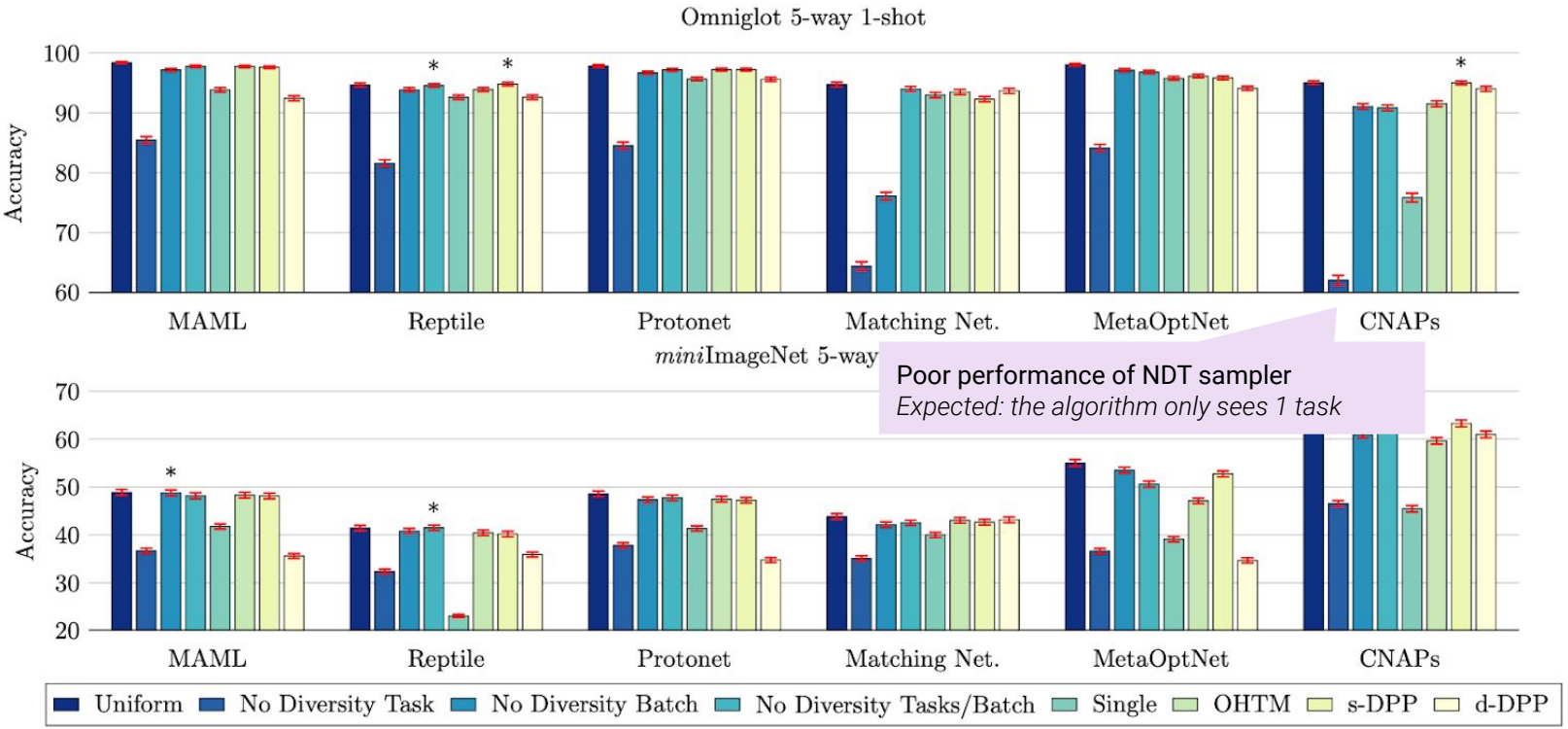
Metric

- MetaOptNet — Hybrid (Metric + Optimization)
- Conditional Neural Adaptive Processes (CNAPS) — Bayesian

Results


Omniglot 5-way 1-shot
Accuracy
100
90
80
70
60
MAML
Reptile
Protonet
Matching Net.
MetaOptNet
CNAPs
*
*
*
miniImageNet 5-way 1-shot
Accuracy
70
60
50
40
30
20
MAML
Reptile
Protonet
Matching Net.
MetaOptNet
CNAPs
*
*
*
Uniform
No Diversity Task
No Diversity Batch
No Diversity Tasks/Batch
Single
OHTM
s-DPP
d-DPP

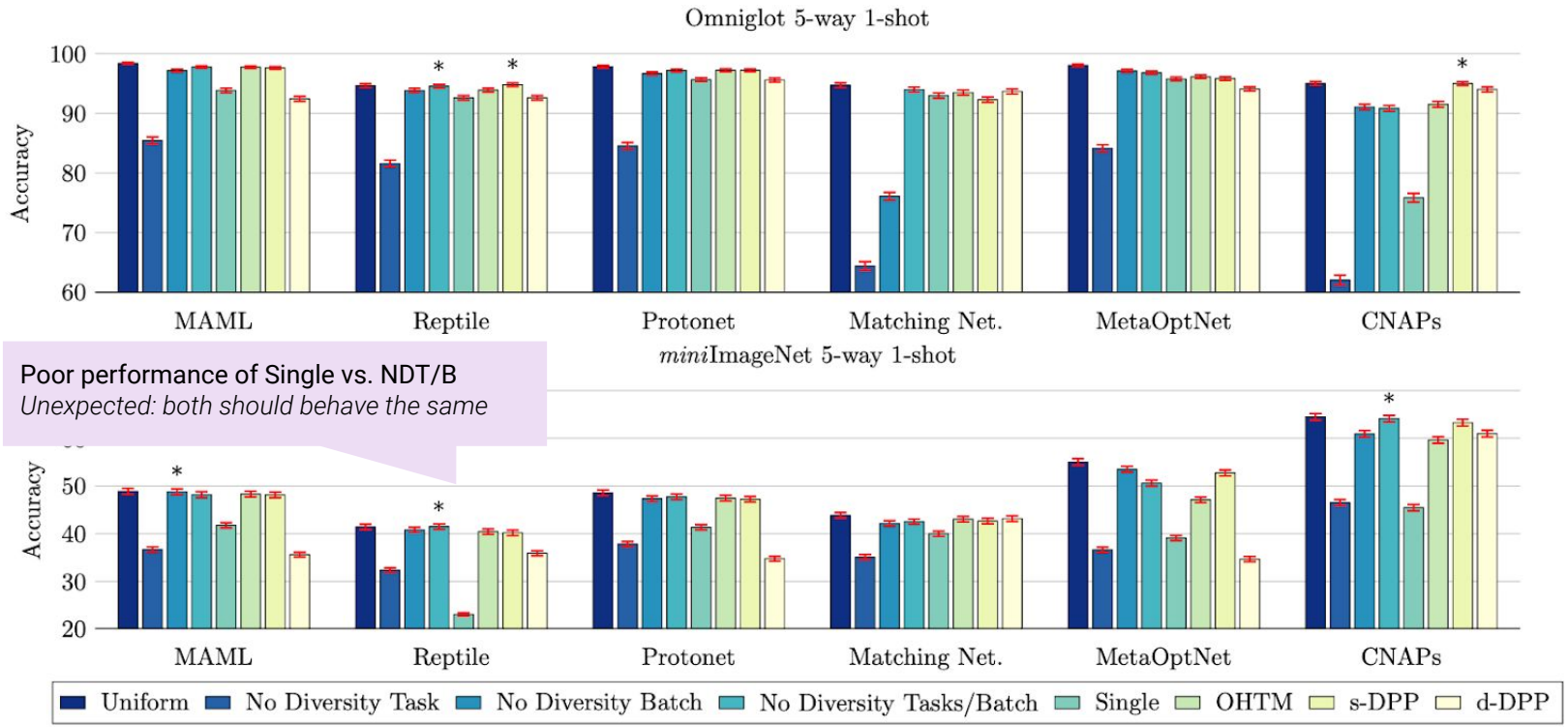
Results



Poor performance of NDT sampler

Expected: the algorithm only sees 1 task

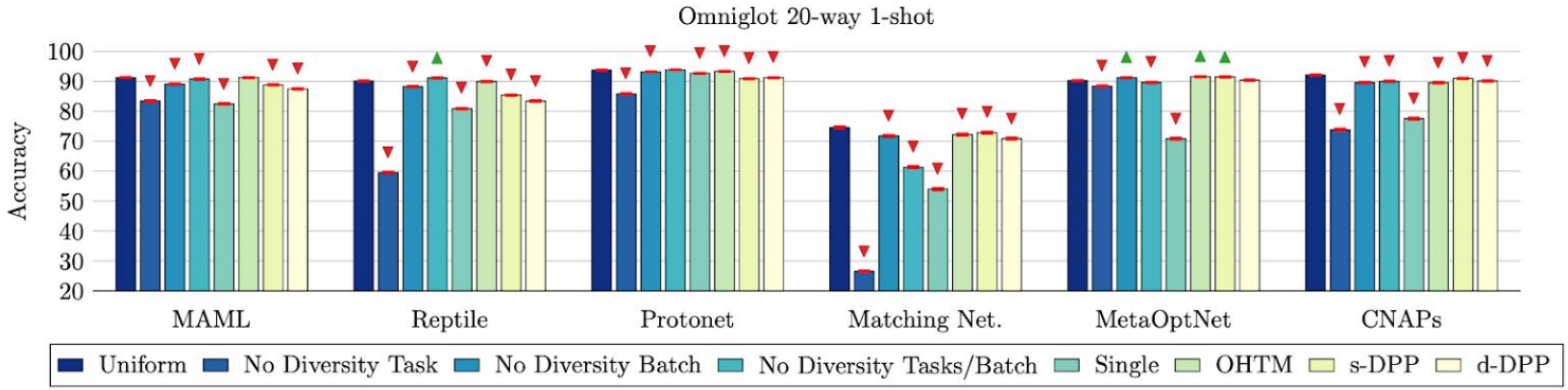
Results



Poor performance of Single vs. NDT/B

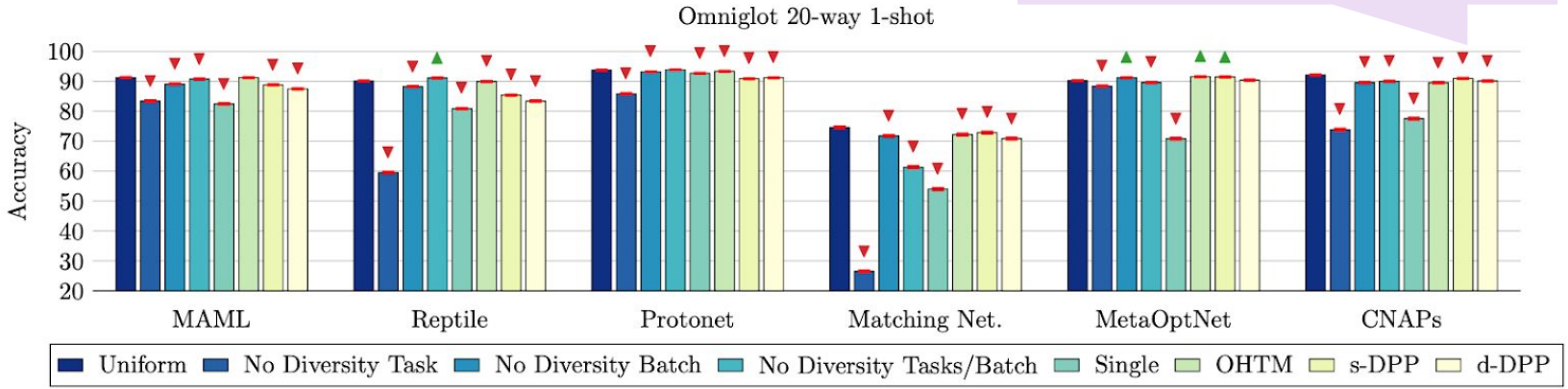
Unexpected: both should behave the same

Results


Omniglot 20-way 1-shot
Accuracy
100
90
80
70
60
50
40
30
20
MAML
Reptile
Protonet
Matching Net.
MetaOptNet
CNAPs
Uniform
No Diversity Task
No Diversity Batch
No Diversity Tasks/Batch
Single
OHTM
s-DPP
d-DPP

Results

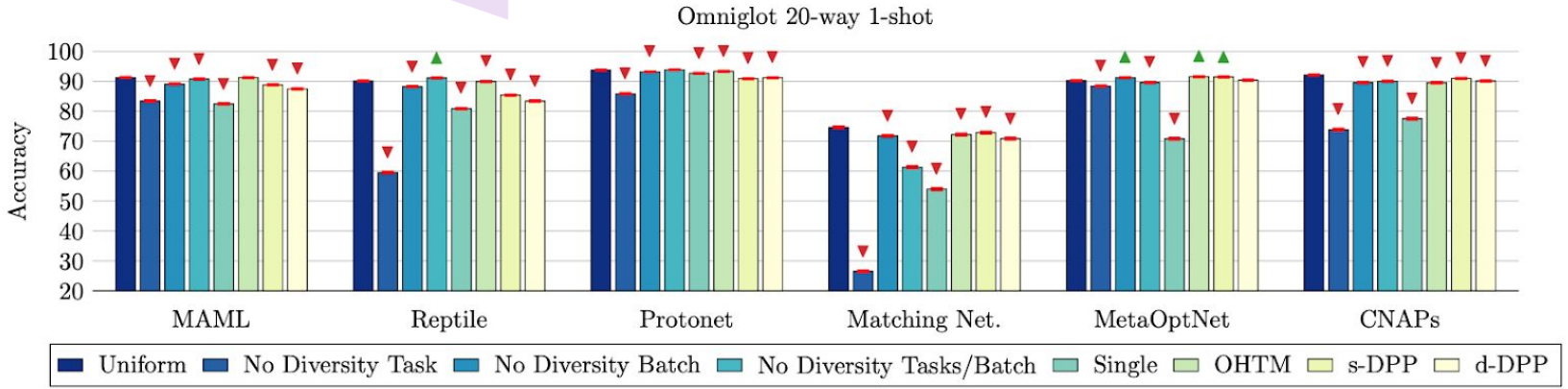
No advantage to have larger diversity
Unexpected: s-DPP & d-DPP usually perform worse



Results

NDTB, NDB and Uniform Sampler have similar performances

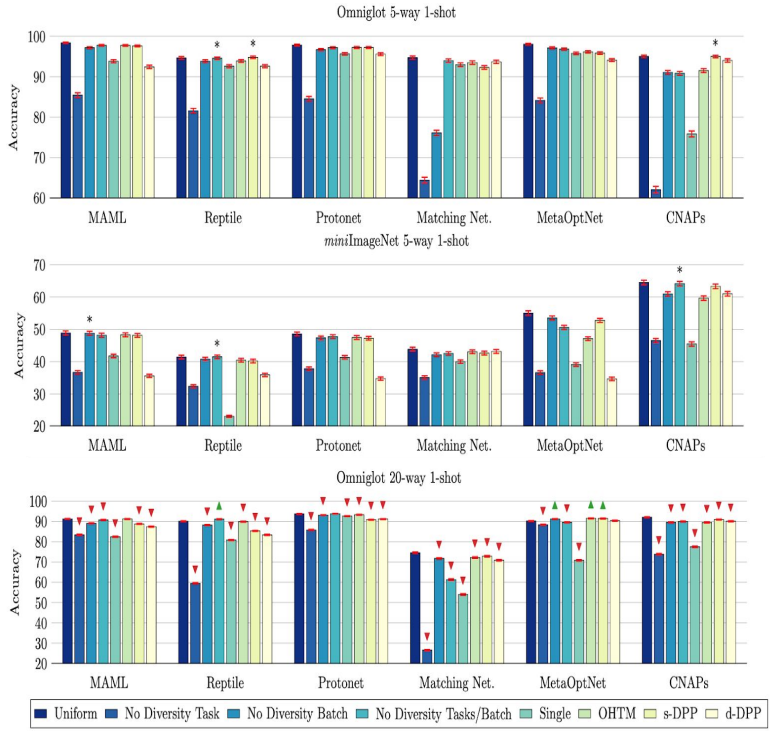
Unexpected: Similar results with only marginal amount of data



And More…!

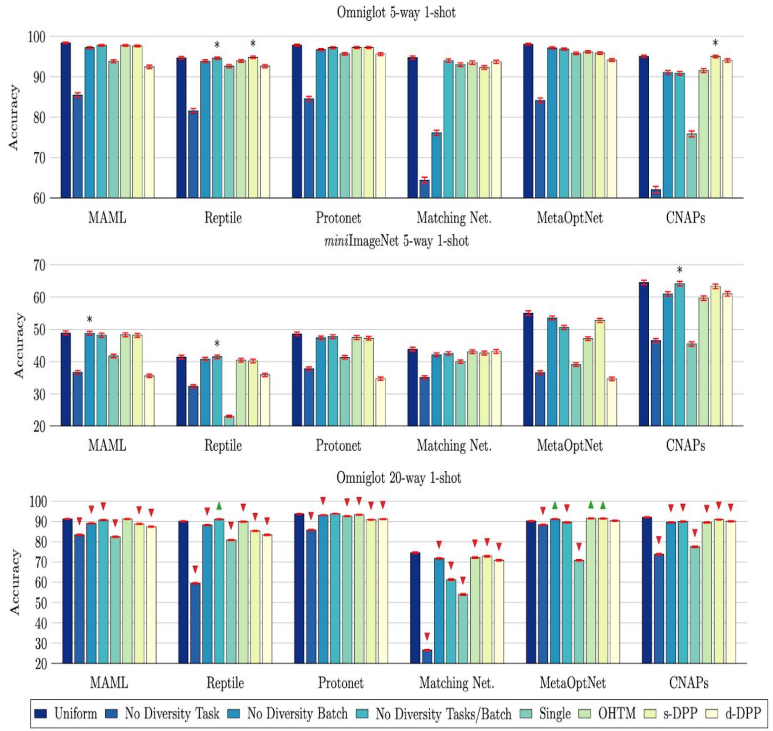
Conclusions

- 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



Conclusions

- **Low task-diversity** (eg. No Diversity Batch) does not hurt performance too much
  - This has impact in practical applications, where uniform sampling is impossible
- **High task-diversity** does not help significantly




Mila
Thanks!