Can we use teacher-student curriculum learning with small teacher networks to enhance meta-learning?

Bertold B. Kovács
B.B.Kovacs-1@student.tudelft.nl

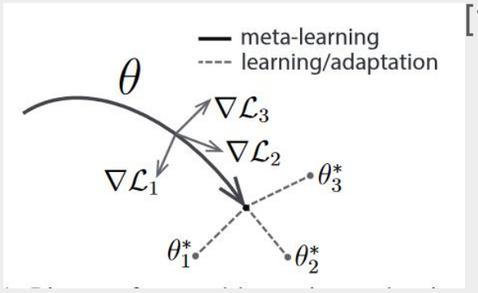
Joery de Vries, supervisor
J.A.deVries@.tudelft.nl

Matthijs Spaan, resp. prof.
M.T.J.Spaan@tudelft.nl

TU Delft

Meta-Learning

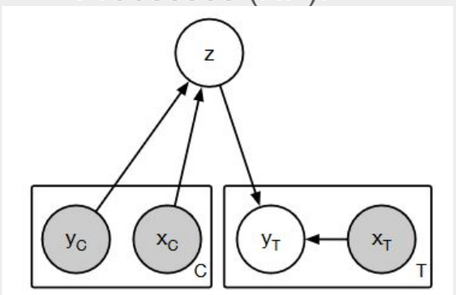
- Goal is to perform well on an unseen task after (minimal) extra samples



[1]

Neural Processes

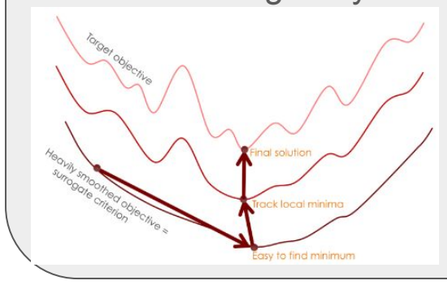
- The algorithm that we use for meta-learning is Neural Processes (NP).



[2]

Curriculum Learning (CL)

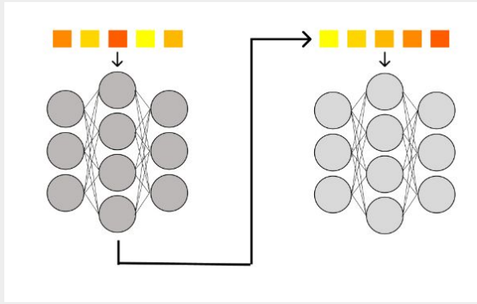
- Easier data points are more useful during early training.

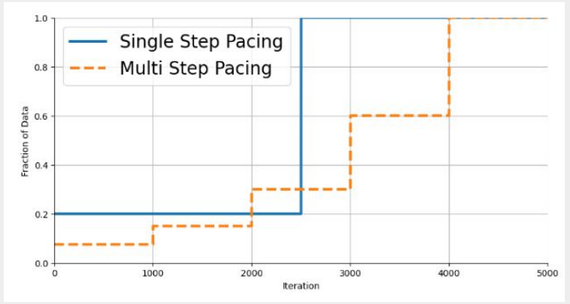


[3]

Teacher-Student Approach

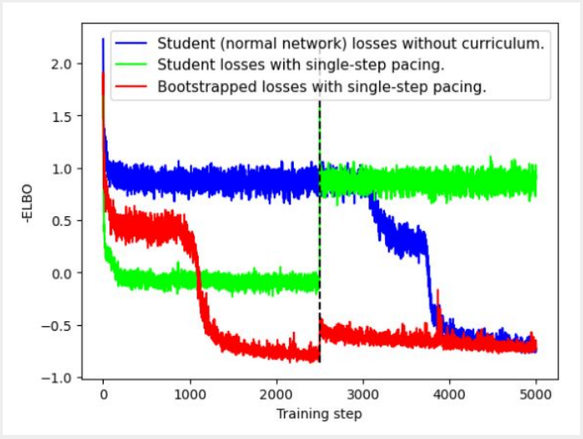
- A CL algorithm needs to answer:
  - Which samples are easy? (scoring) => *we can train a "teacher" without CL*
  - What portion should we sample from? (pacing)

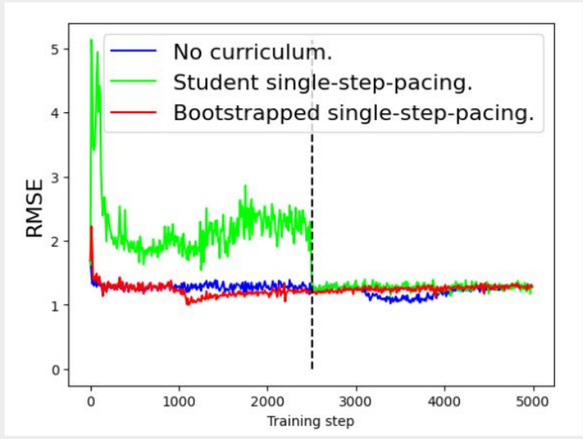


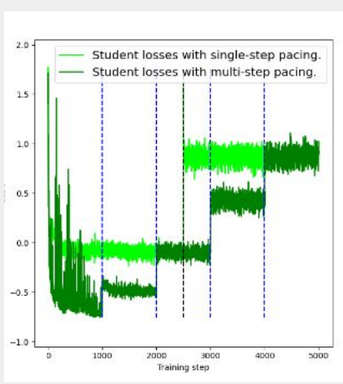


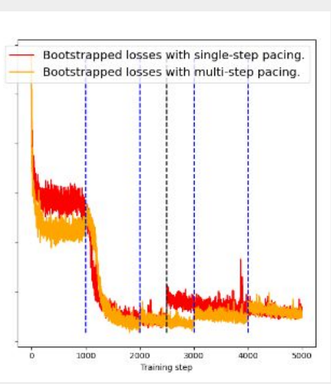
- Training an extra network is expensive!
  - Solutions: complex network pretrained for a different purpose (transfer teacher), or training the same network on the same dataset (bootstrap CL).
  - **Can we go further, and train an even smaller network than the student?**

Results









Evaluation

- Bootstrap CL performs well.
- The teacher smaller than the student performs worse than no CL
- Unclear if this holds for other architectures, domains.

References

[1] Chelsea Finn, Pieter Abbeel, and Sergey Levine. "Model-agnostic meta-learning for fast adaptation of deep networks". In: International conference on machine learning. PMLR. 2017, pp. 1126–1135
[2] Marta Garnelo et al. "Neural processes". In: arXiv preprint arXiv:1807.01622 (2018)
[3] Xin Wang, Yudong Chen, and Wenwu Zhu. "A survey on curriculum learning". In: IEEE transactions on pattern analysis and machine intelligence 44.9 (2021), pp. 4555–4576