On the Efficacy of Differentially Private Few-shot Image Classification

Workshop on the pitfalls of limited data and computation for Trustworthy ML, ICLR 2023

Marlon Tobaben*[1], Aliaksandra Shysheya*[2], John Bronskill[2], Andrew Paverd[3],

Shruti Tople[3], Santiago Zanella-Beguelin[3], Richard E. Turner[2], Antti Honkela[1]

[1] UNIVERSITY OF HELSINKI

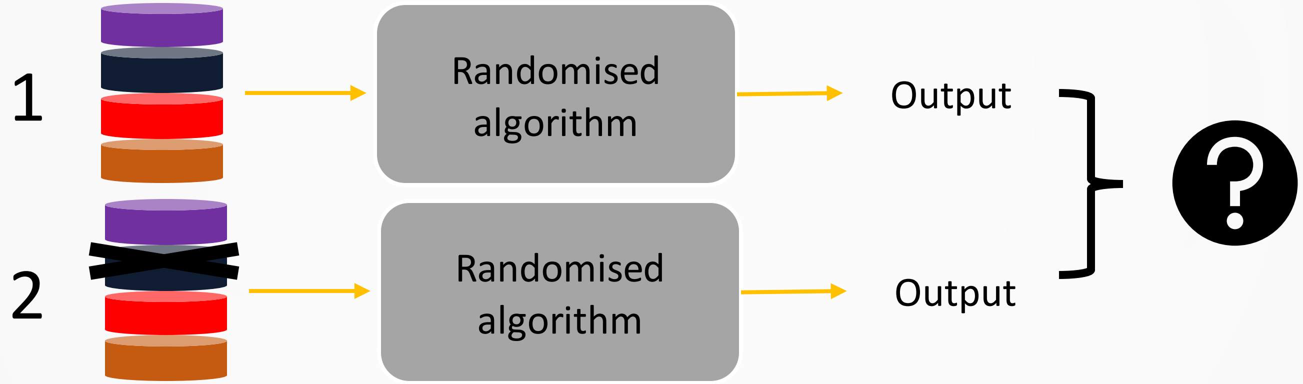
[2] UNIVERSITY OF CAMBRIDGE

[3] Microsoft

Paper: arXiv:2302.01190, Code: https://github.com/cambridge-mlg/dp-few-shot

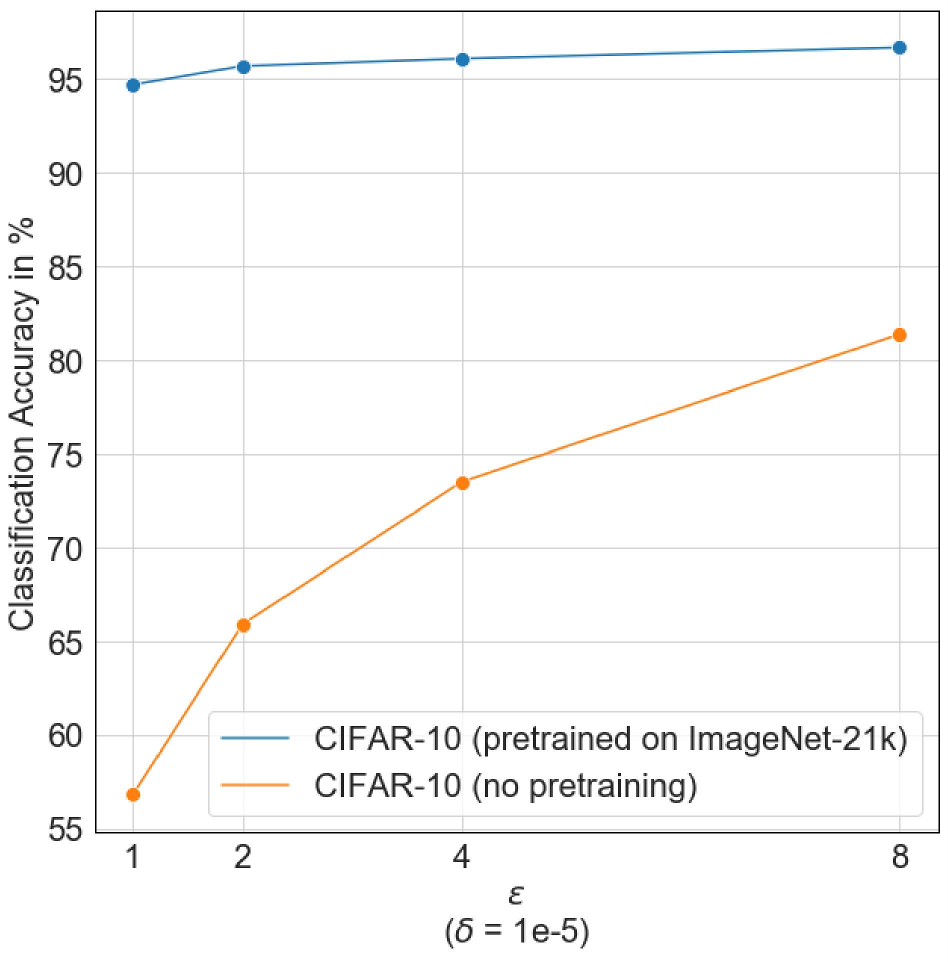
Differential Privacy (DP)

- The gold standard for formalizing privacy guarantees
- Looking at the output, can't tell if a data point was in the dataset or not



- (ε, δ)-DP with privacy budget ε ≥ 0 (lower means more private) and additive error δ ∈ [0, 1] bounds how much the output distribution can diverge on adjacent datasets

SOTA DP deep learning relies on transfer learning



Assumptions:

- use backbones pretrained on large public datasets
- downstream data is private

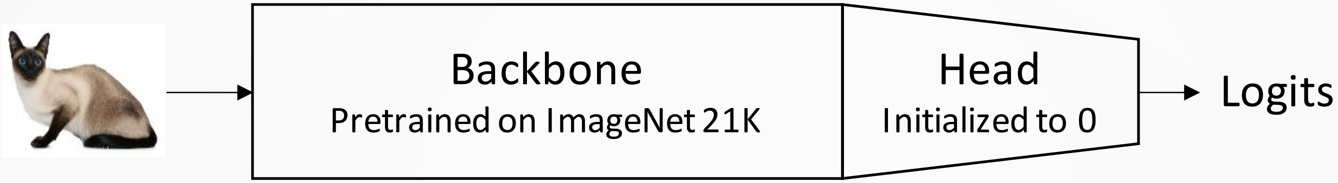
Previous work focuses on downstream datasets that are:

- large
- very similar to the pretraining dataset

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

De, S., Berrada, L., Hayes, J., Smith, S. L., & Balle, B. (2022). Unlocking high-accuracy differentially private image classification through scale. arXiv:2204.13650.
Tramèr, F., Kamath, G., & Carlini, N. (2022). Considerations for Differentially Private Learning with Large-Scale Public Pretraining. arXiv:2212.06470.

Image Classification - Transfer Learning



Backbone
Pretrained on ImageNet 21K

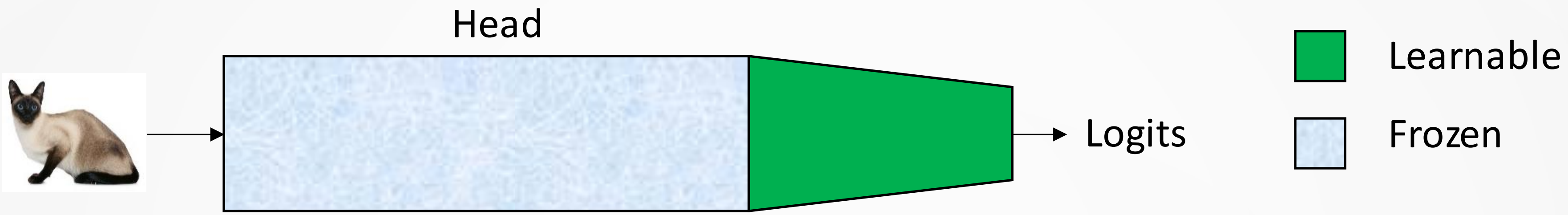
Head
Initialized to 0

Logits

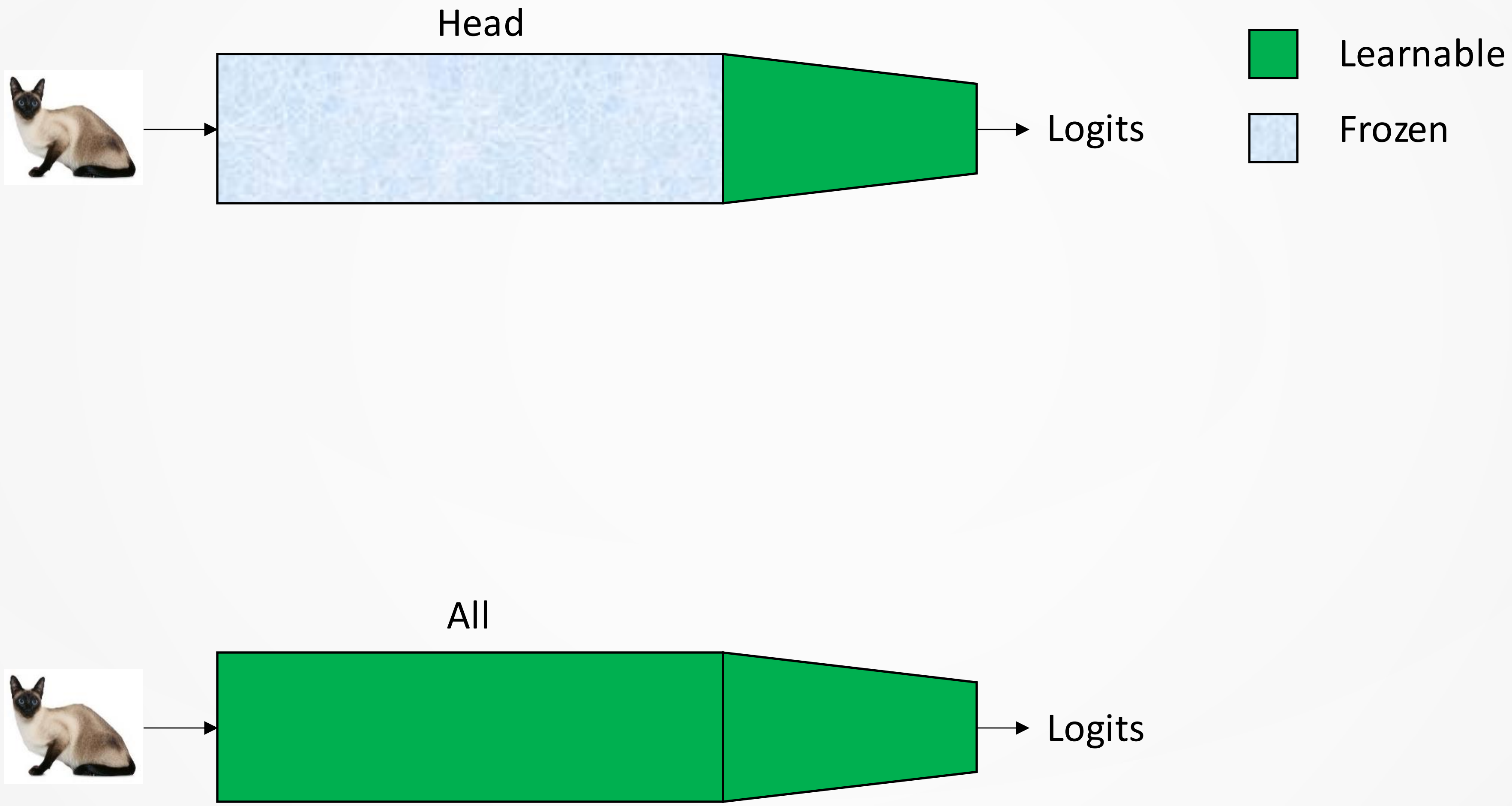
- BiT-M-R50x1 (R-50)
  (23.5M parameters)
- Vision Transformer
  VIT-Base-16 (VIT-B)
  (85.8M parameters)

Linear Layer

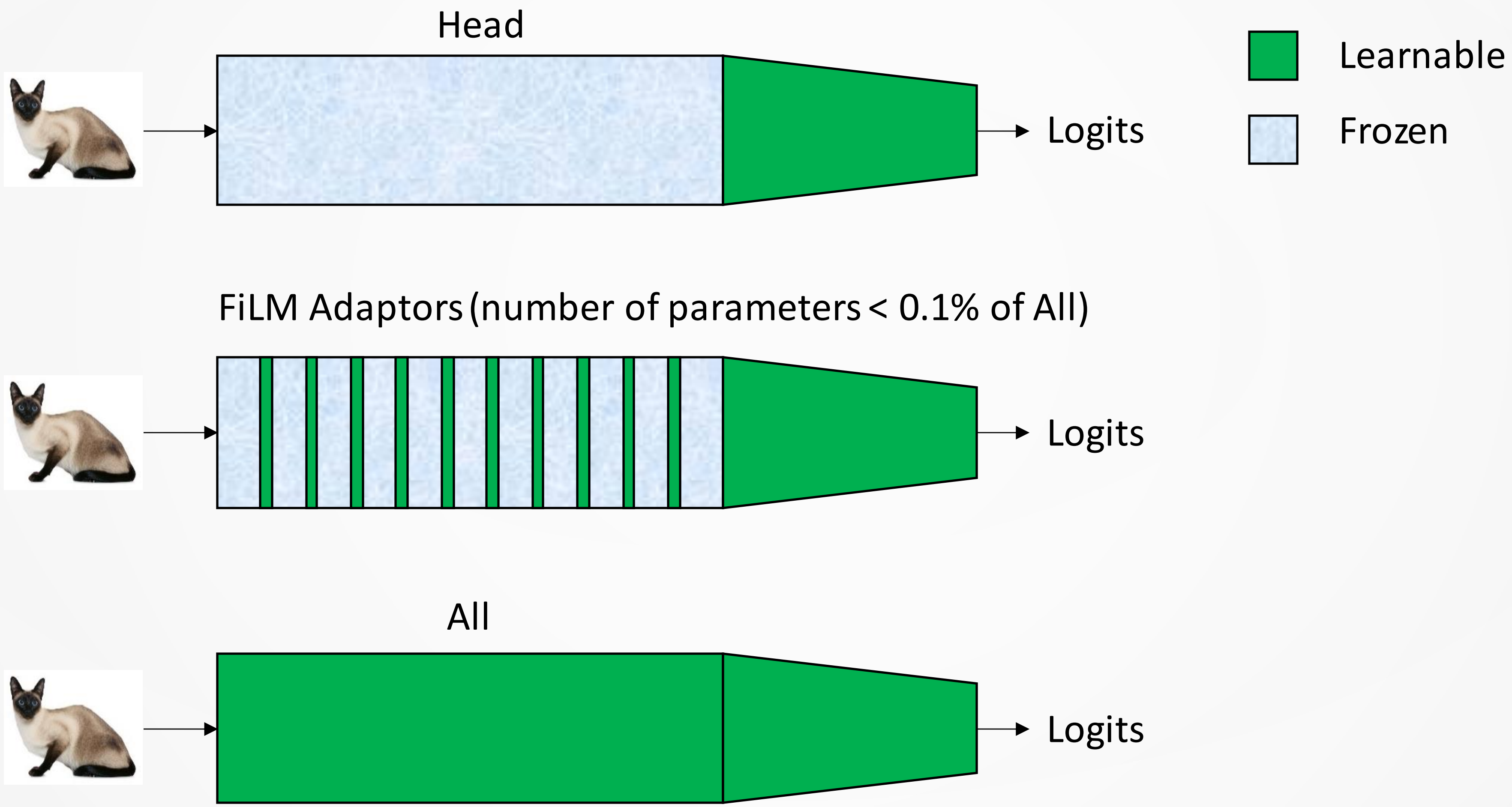
Which parameters are fine-tuned?



Which parameters are fine-tuned?



Which parameters are fine-tuned?



Head

Logits

🟩 Learnable

▨ Frozen

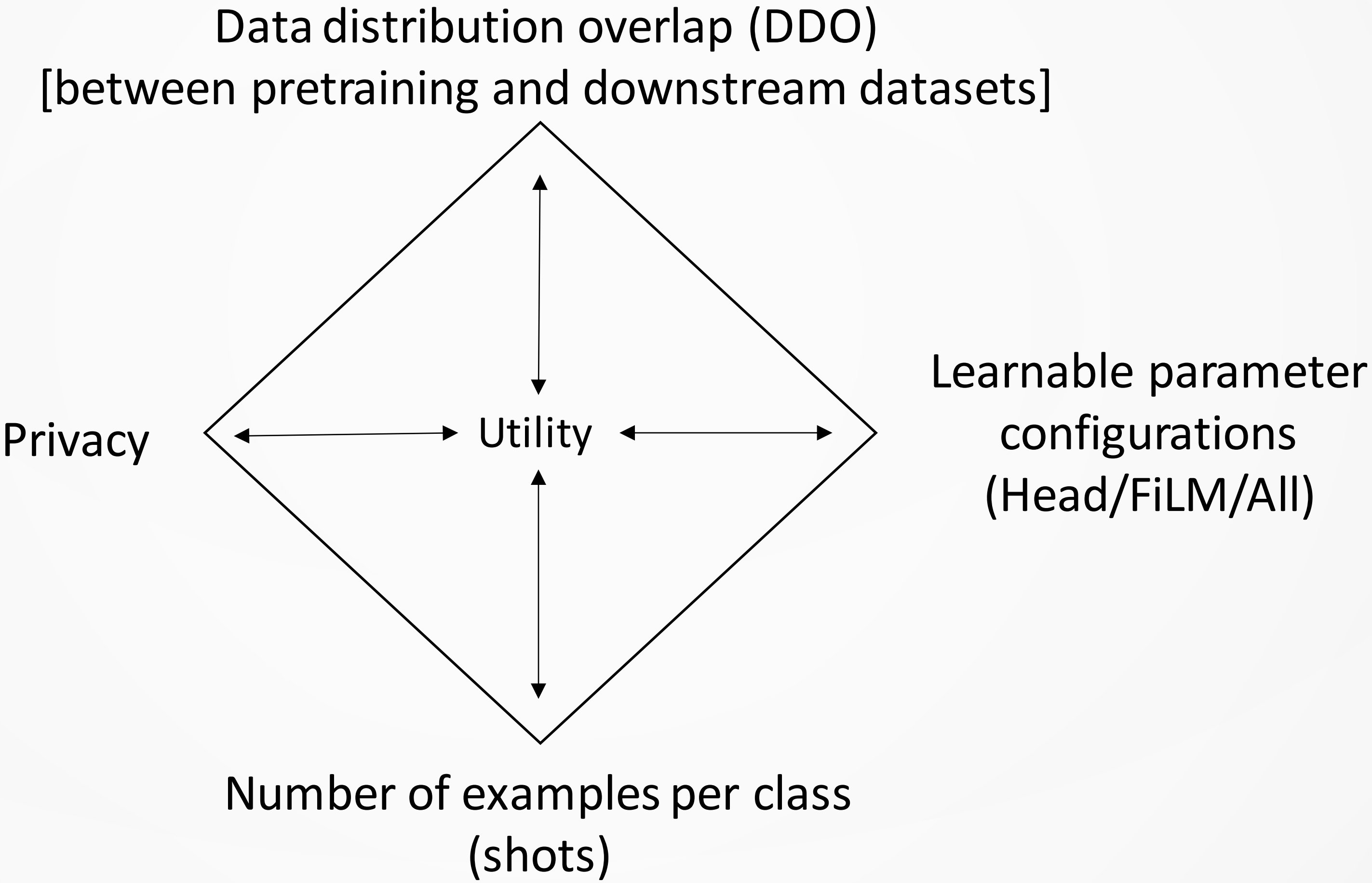
FiLM Adaptors (number of parameters < 0.1% of All)

Logits

All

Logits

Tradeoffs in this work



Data distribution overlap (DDO)
[between pretraining and downstream datasets]

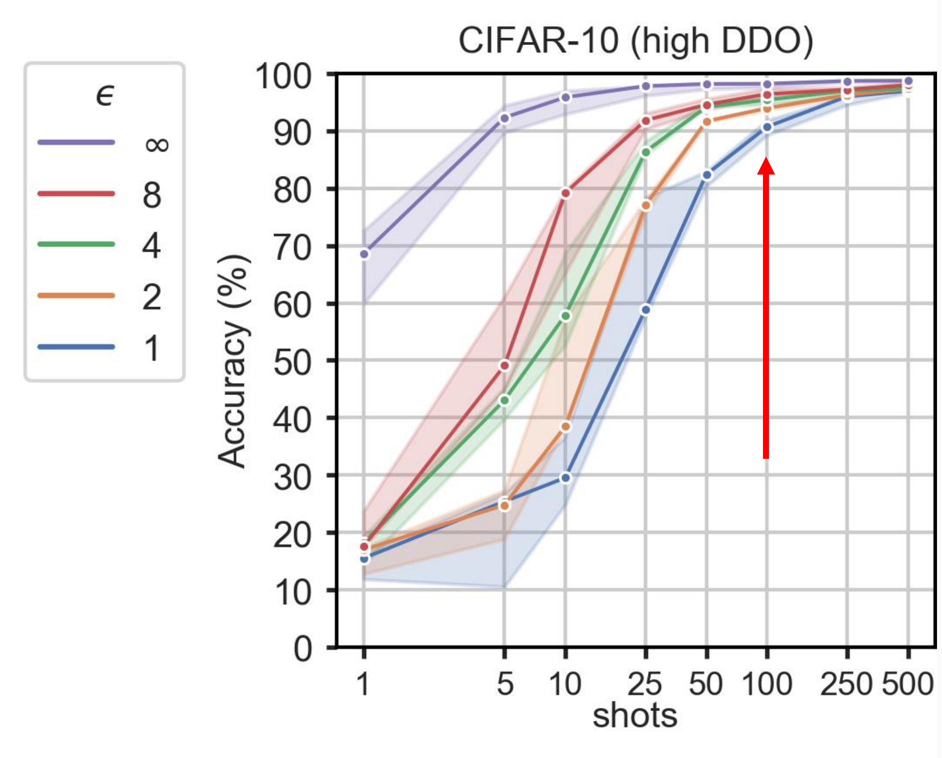
Privacy

Utility

Learnable parameter
configurations
(Head/FiLM/All)

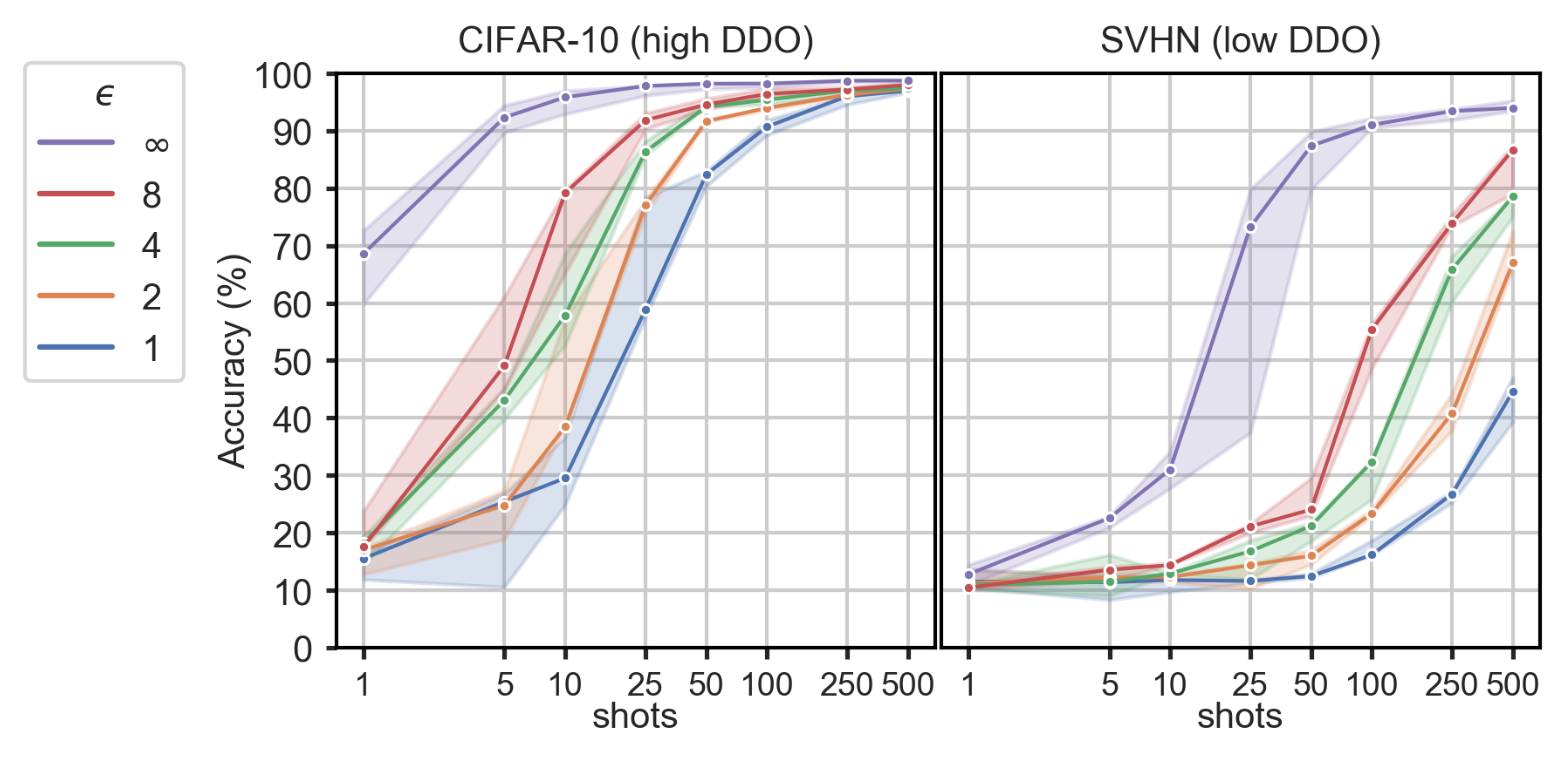
Number of examples per class
(shots)

Effect of Shots and Privacy



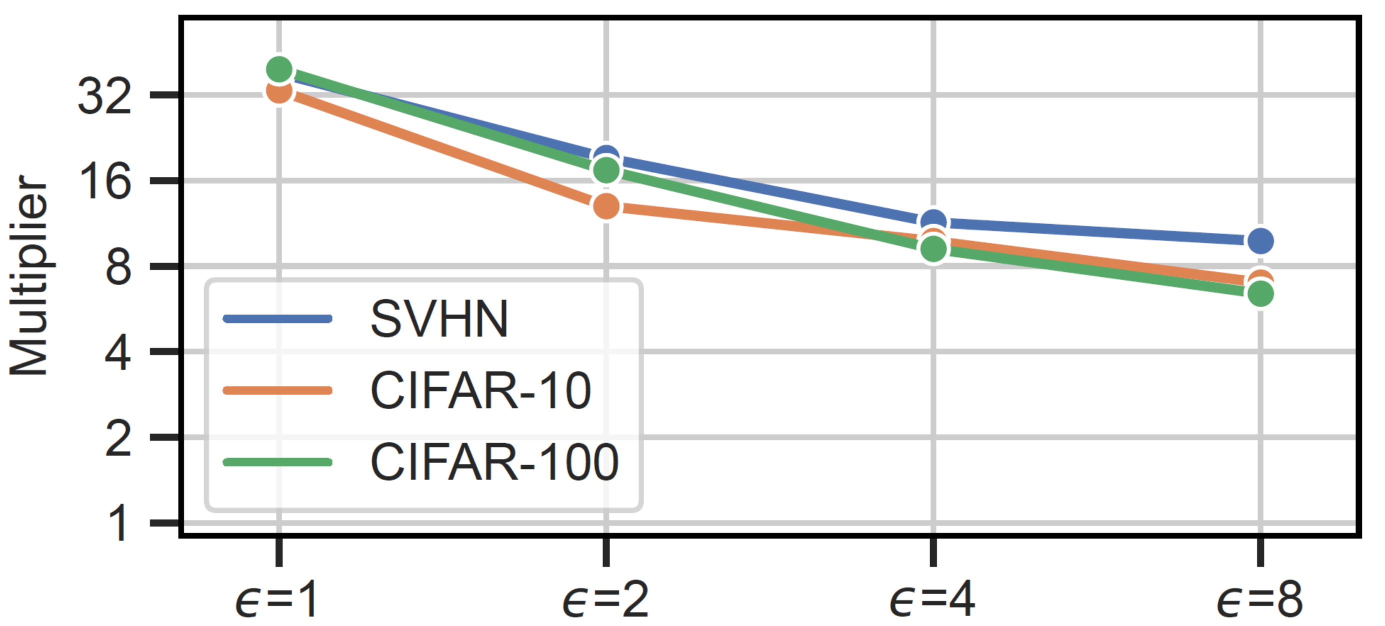
- At low shot, accuracy degrades significantly with increasing privacy level
- High DDO: 100 shots are required for high accuracy (90%)

Effect of Shots and Privacy



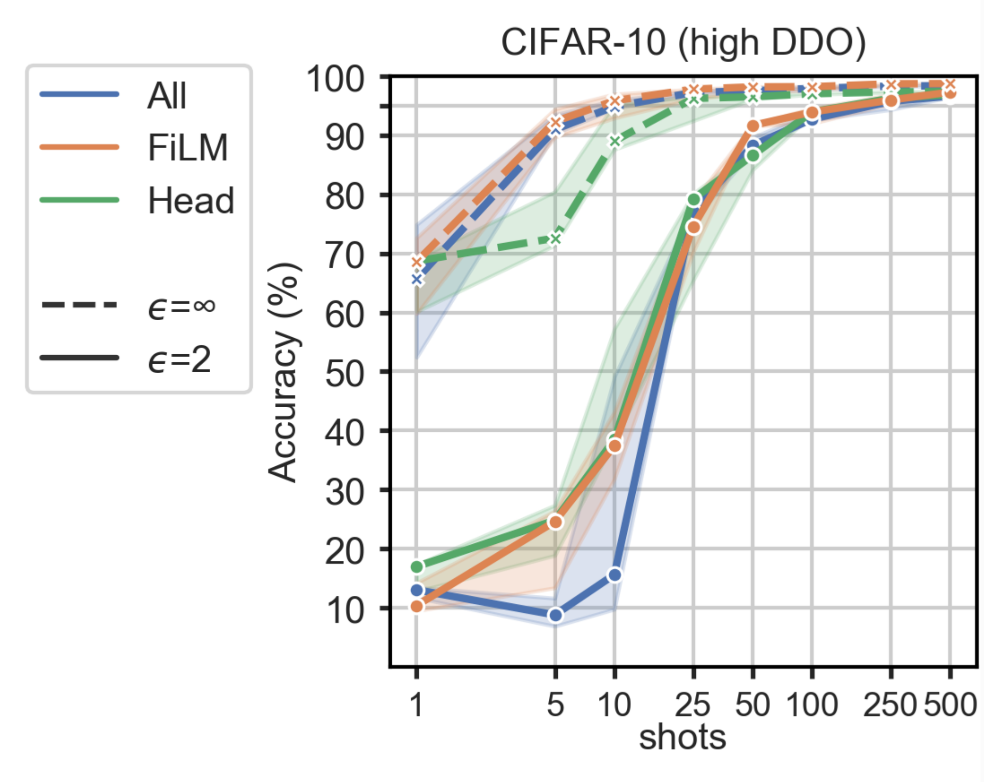
- At low shot, accuracy degrades significantly with increasing privacy level
- High DDO: 100 shots are required for high accuracy (90%)
- Low DDO: More data is required to close gap to non-private performance

How much data is required to match non-private accuracy?



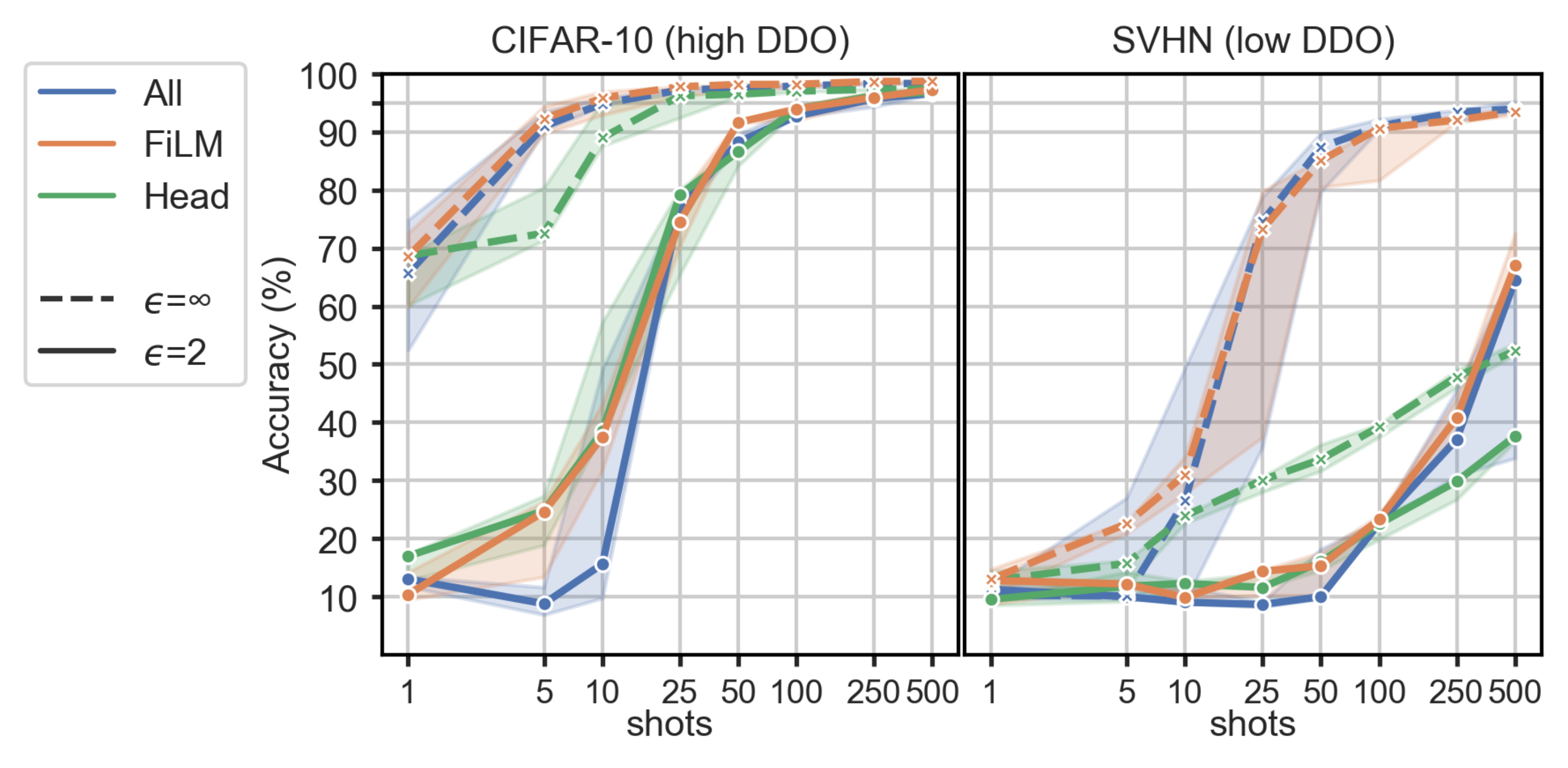
- 32x data required at ε=1 to match non-private accuracy when shots = 5

Comparing different configurations



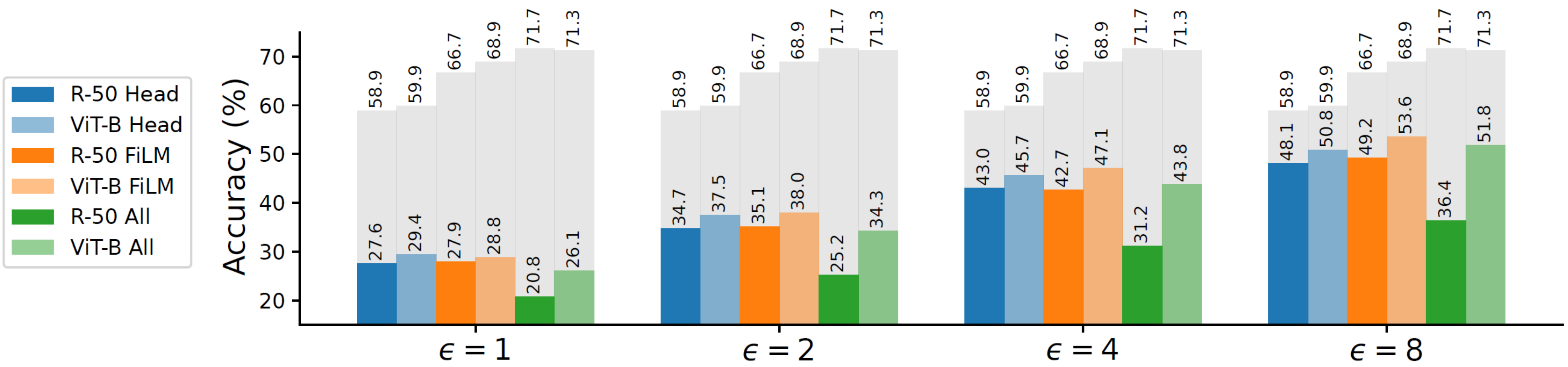
- In General: FiLM is at least as good or better as All and Head

Comparing different configurations



- In General: FiLM is at least as good or better as All and Head
- Low DDO: Head falls short

Average accuracy of 19 diverse VTAB datasets



- Non-private: All > FiLM > Head (Gray)
- Private FiLM >= Head > All (colored bars)

Thanks for listening

- Paper: arXiv:2302.01190
- Code: https://github.com/cambridge-mlg/dp-few-shot

- Check out our posters today:
  - This work:
    On the Efficacy of Differentially Private Few-shot Image Classification
  - Learnings applied to federated learning:
    Differentially Private Federated Few-shot Image Classification