UNIVERSITY OF CAMBRIDGE

Microsoft

LITE:
Memory Efficient Meta-Learning with Large Images

John Bronskill[1]

Daniela Massiceti[2]

Massimiliano Patacchiola[1]

Katja Hofmann[2]

Sebastian Nowozin[2]

Richard E. Turner[1]

[1]University of Cambridge

[2]Microsoft Research

NeurIPS 2021

aka.ms/lite-neurips2021

TL;DR

The Problem

Meta-learning algorithms for few-shot classification are constrained to small images (e.g. 84 x 84 pixels) during training due to memory constraints.

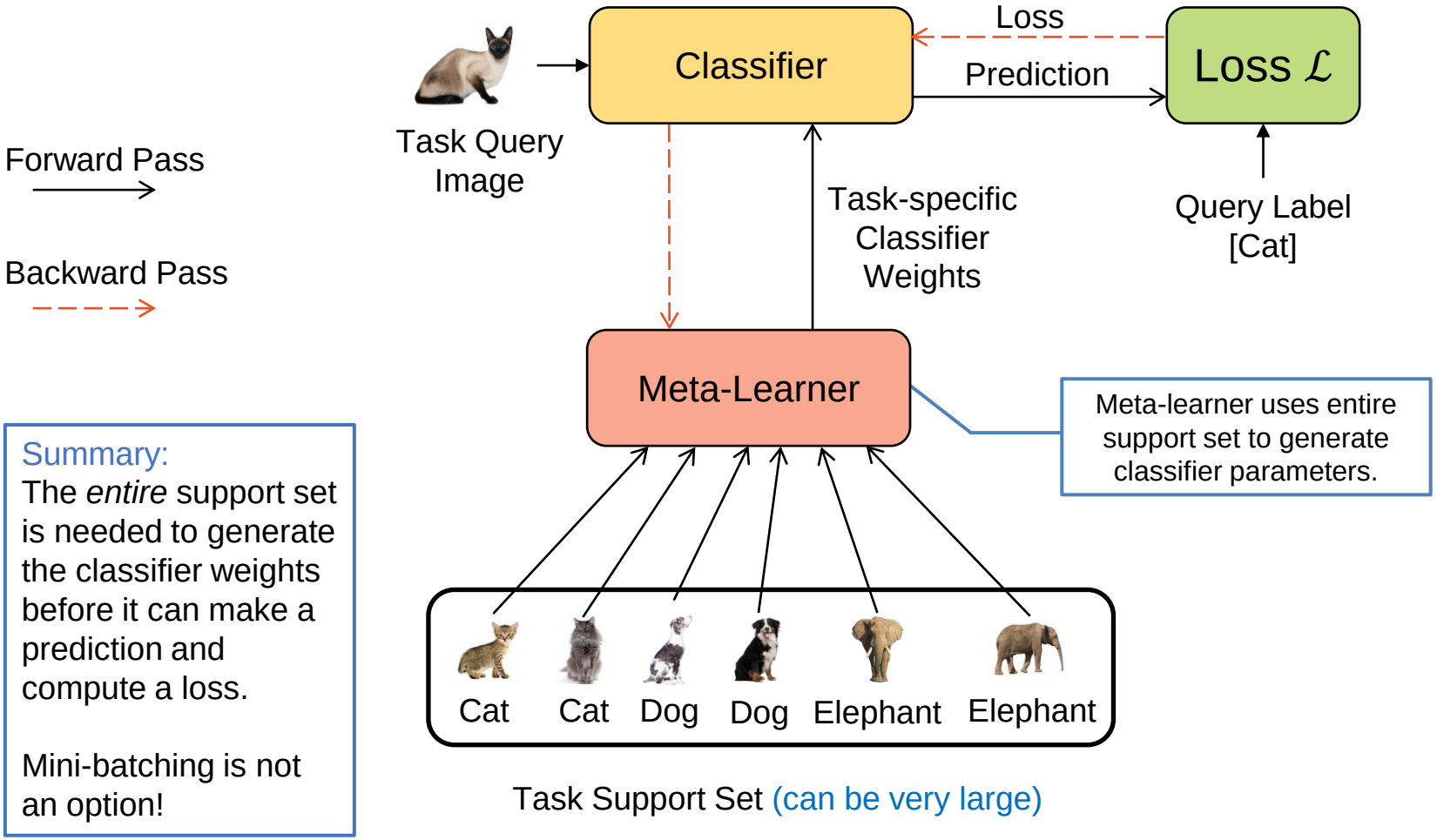
The Solution

LITE, which reduces training memory usage by back-propagating only a small random subset of a task's examples.

The Benefits

1. LITE approximation is an unbiased estimate of the true gradient.
2. Greater than 12% classification accuracy gains from using larger images on a single GPU.
3. SOTA results for meta-learners on challenging VTAB+MD and ORBIT benchmarks.

Problem:
GPU memory limits constrain meta-learning algorithms



Summary:
The *entire* support set is needed to generate the classifier weights before it can make a prediction and compute a loss.

Mini-batching is not an option!

Current Solutions

Solution	Issue
1. Small (e.g. 84 x 84) images	Classification accuracy suffers (>10% drop[1]); can't use 224x224 pretrained models.
2. Small (subsampled) tasks	Low performance on large tasks.
3. Multiple GPUs (model parallelism)	May not be available; difficult to code.
4. Checkpointing (recompute activations on demand)	Slow; not enough memory savings.

[1] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L.-C. *Mobilenetv2: Inverted residuals and linear bottlenecks*. CVPR, 2018.

Solution:
LITE (Large Image and Task Episodic Training)

Forward Pass

Forward Pass, Disabled Gradient

Backward Pass

Task Query Image

Classifier

Loss

Prediction

Loss $\mathcal{L}$

Query Label [Cat]

Task-specific Classifier Weights

Meta-Learner

Entire support set is passed forward through the meta-learner, but only a small subset (<10%) have gradients enabled saving memory to store activations.

Gradients: On

Gradients: Off

Cat

Cat

Dog

Dog

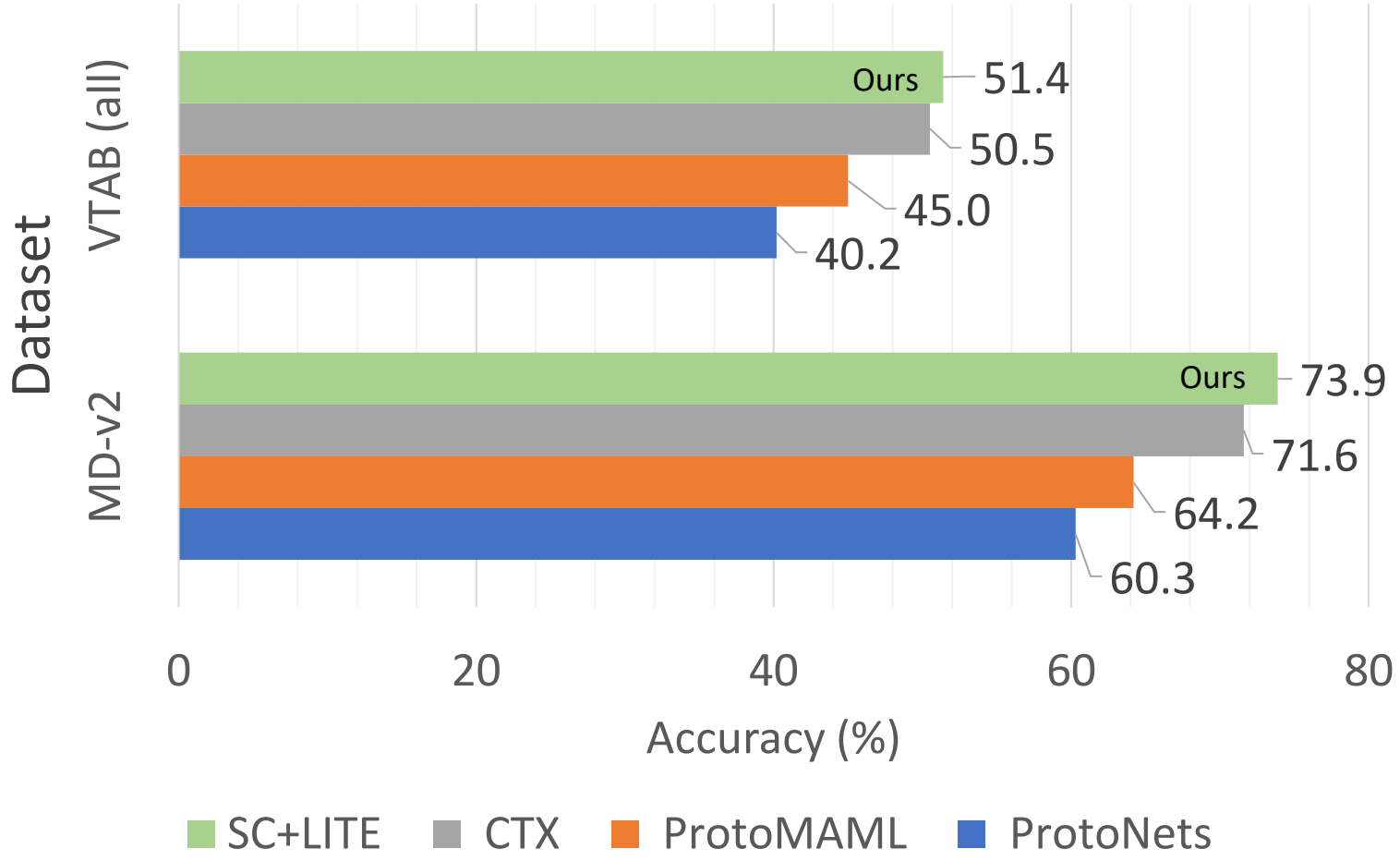
Elephant

Elephant

Task Support Set (can be very large)

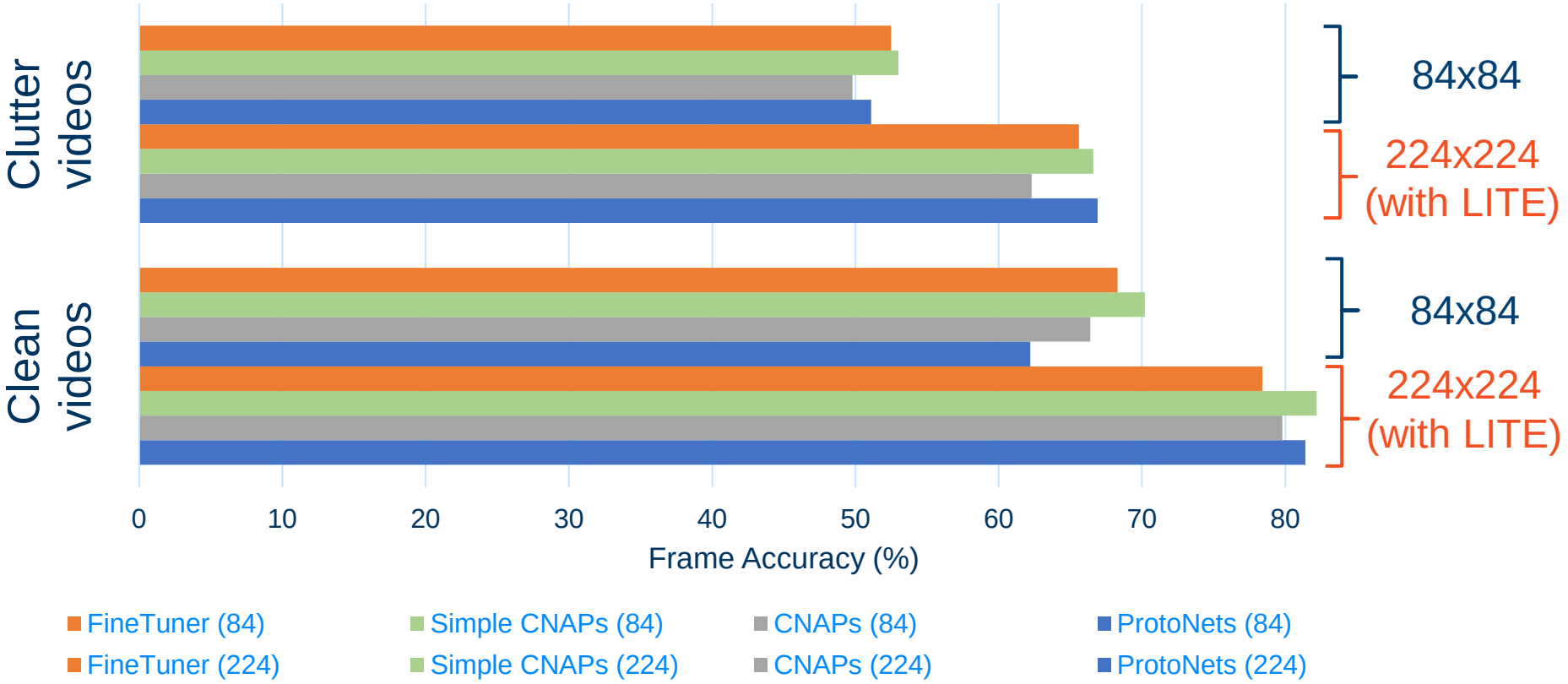
Note:
LITE is an approximation, but it is an unbiased estimate of the true gradient!

LITE is SOTA among Meta-Learners on VTAB+MD[1]



[1] Dumoulin, Vincent, et al. "Comparing Transfer and Meta Learning Approaches on a Unified Few-Shot Classification Benchmark." arXiv preprint arXiv:2104.02638 (2021).

LITE is SOTA on ORBIT



Thanks for watching!

Paper

https://openreview.net/forum?id=x2pF7Tt_S5u

Code

VTAB+MD: https://github.com/cambridge-mlg/LITE

ORBIT: aka.ms/orbit-code