Faster DNN Training With Selective Backpropagation

Angela Jiang, Daniel Wong, Michael Kaminsky†, Michael A. Kozuch†, Padmanabhan Pillai†,
David G. Andersen, Gregory R. Ganger

Overview

Can we speed up DNN training by backpropogating only useful examples?

Motivation

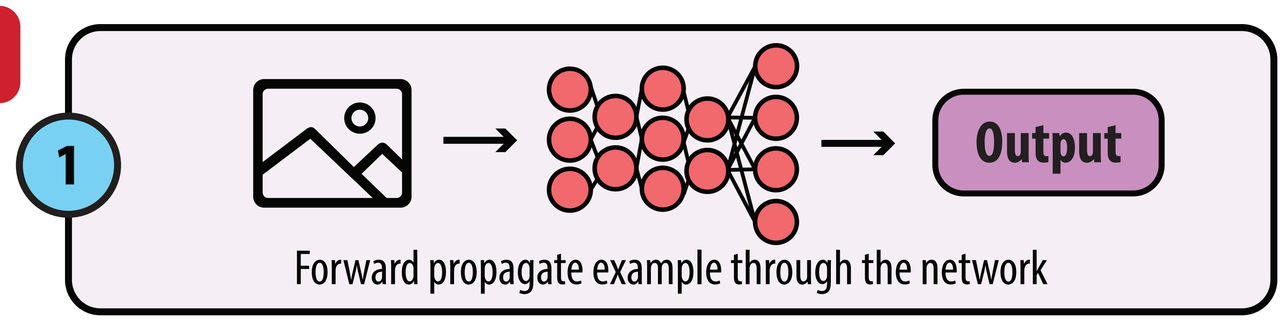
- Labeled datasets are getting larger
- Often not enough time or resources to train on whole dataset (e.g., ImageNet)
- Fast inference (e.g., with accelerators) => Training bottlenecked by backprop

Goal

- Speed up training by reducing the number of backprops
  - Learn mostly from surprising examples that have a lot to teach the network

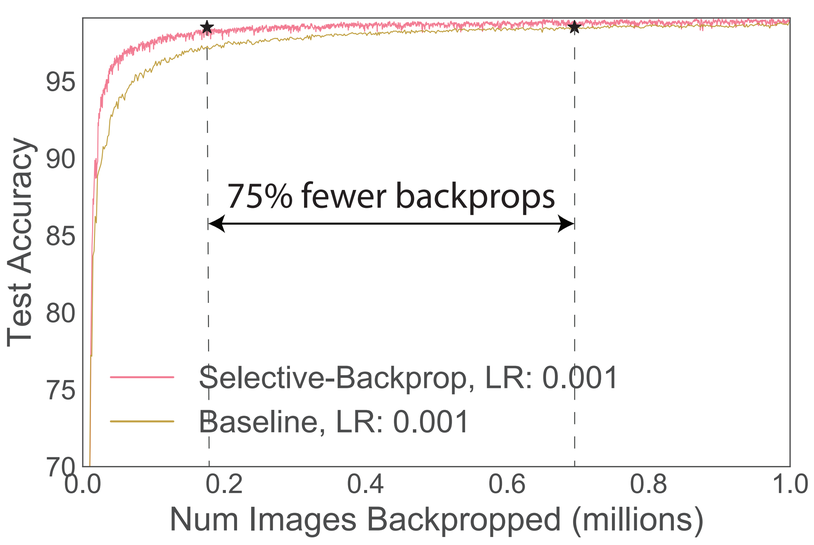
Approach

- Identify useful examples using inference (the output of the forward pass)
  - If example's output is different from target, learn from this example
- Changes sampling distribution of dataset, biased towards interesting examples



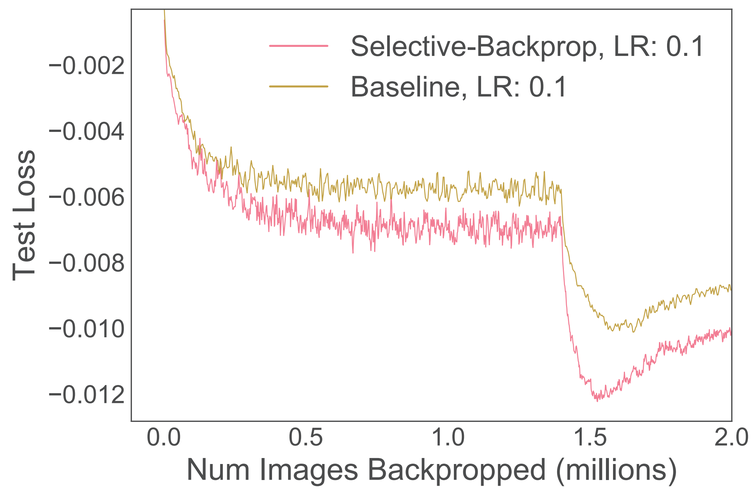
Training wth Selective-Backprop (SB)

MNIST



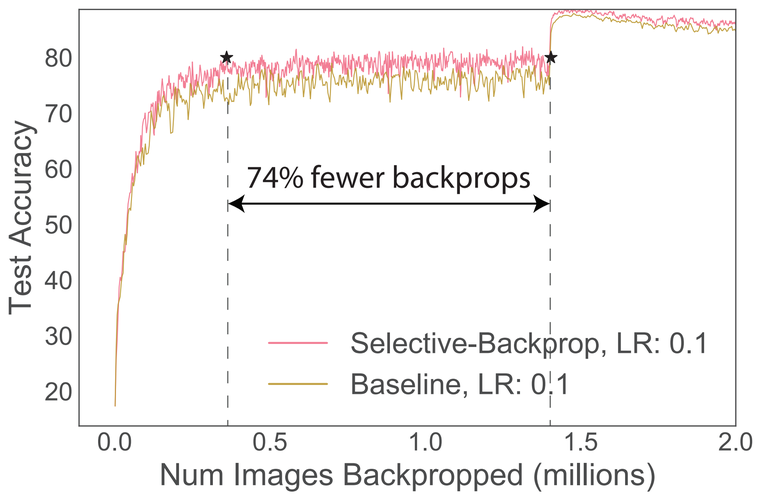
- Baseline does not filter examples
- SB filters >80% of examples of MNIST
- SB gets 98.5% acc. with 75% fewer backprops

CIFAR10



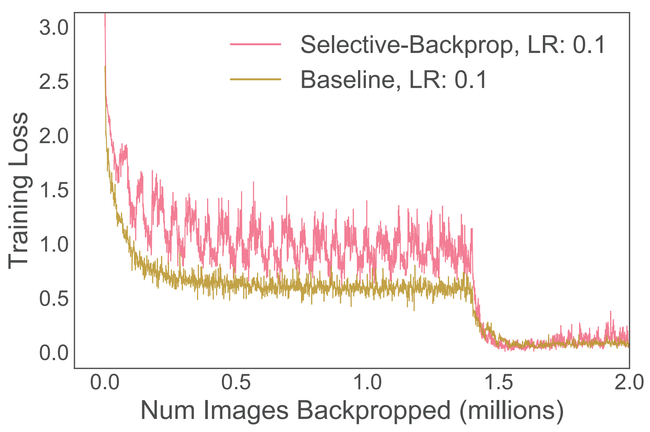
- SB filters >75% of CIFAR10 examples
- SB reduces test loss with fewer examples

CIFAR10



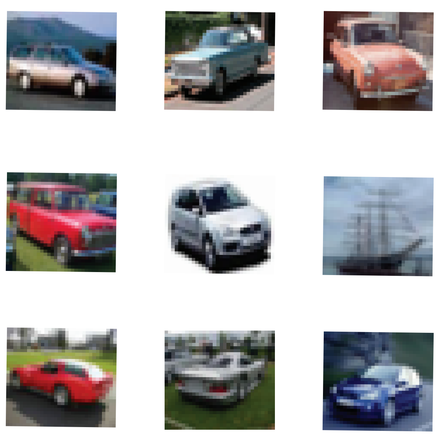
- SB gets 80% acc. with 74% fewer backprops
- Acc has sharp increase from learning rate scheduler

Diving into CIFAR10

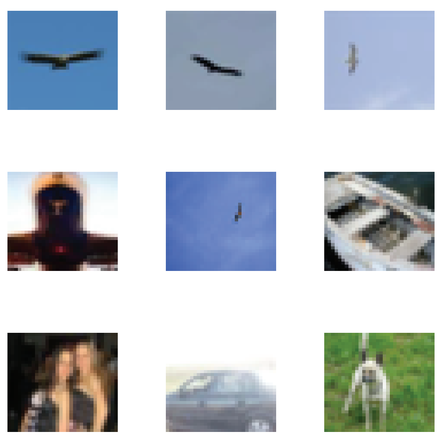


- SB trains images with higher loss

Easy Examples



Hard Examples



Next steps

- Use SB to reduce wall-clock time
- Use SB on high frame-rate video datasets
  - To avoid filtering useful frames
- Use SB offline to dictate order to train examples
  - To improve training of other models