Faster and More Accurate DNN Training With Selective Backpropagation

Angela Jiang, Daniel Wong, Giuio Zhou, Dave Andersen, Jeff Dean*, Greg Ganger, Michael Kaminsky†, Michael A. Kozuch†, Padmanabhan Pillai†
Carnegie Mellon University; * Google AI; †Intel Labs

Overview

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

Motivation

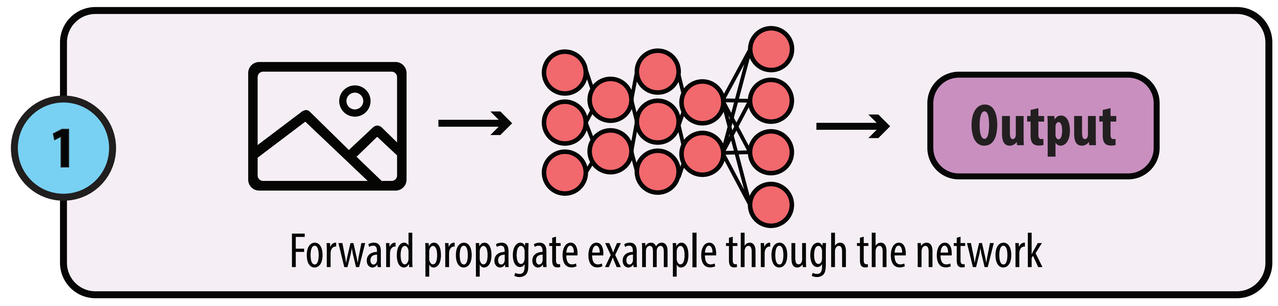
- Labeled datasets are getting larger
- Not enough time/resources to train on whole dataset (e.g., ImageNet)
- Accelerated inference (w/ TPUs) => Training bottlenecked by backprop

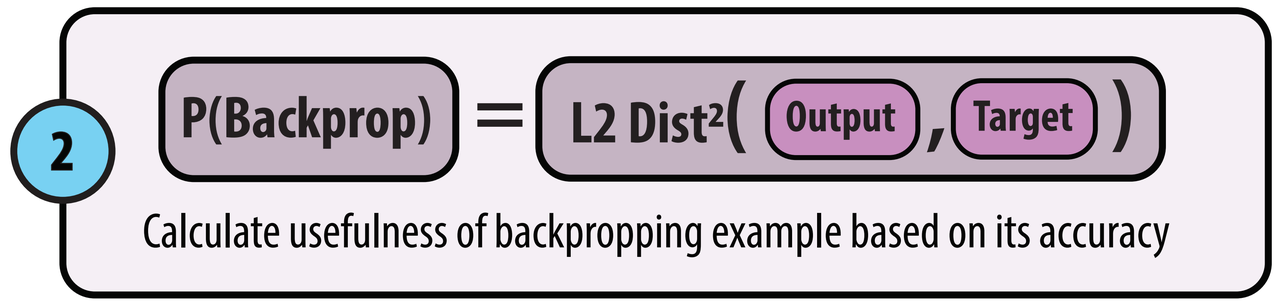
Goal

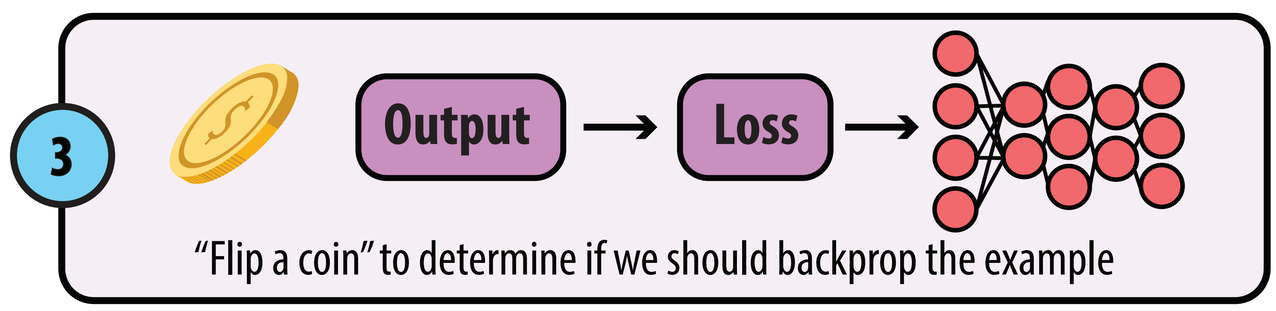
- Speed up training by reducing the number of backprops
  - Learn from surprising examples that have more to teach the network

Approach

- Identify useful examples using inference (the output of the forward pass)
  - If example's output is different from target, learn from this example

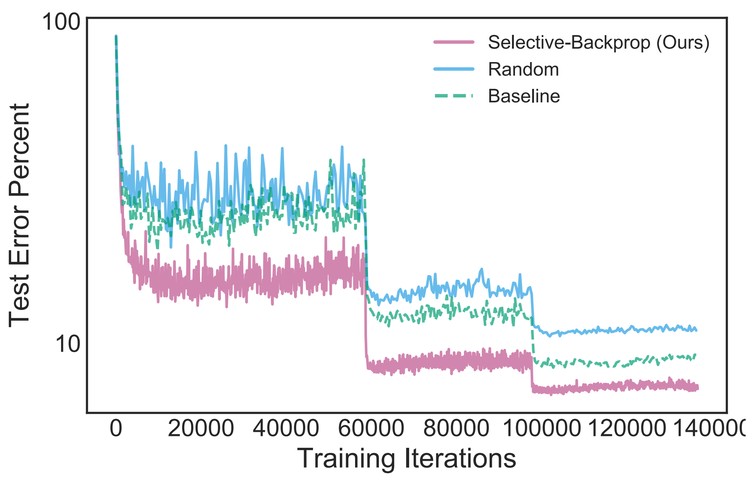






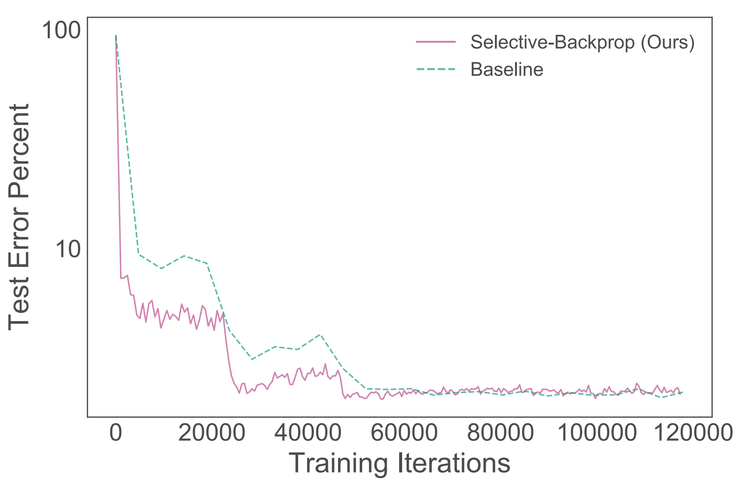
Training wth Selective-Backprop (SB)

CIFAR10



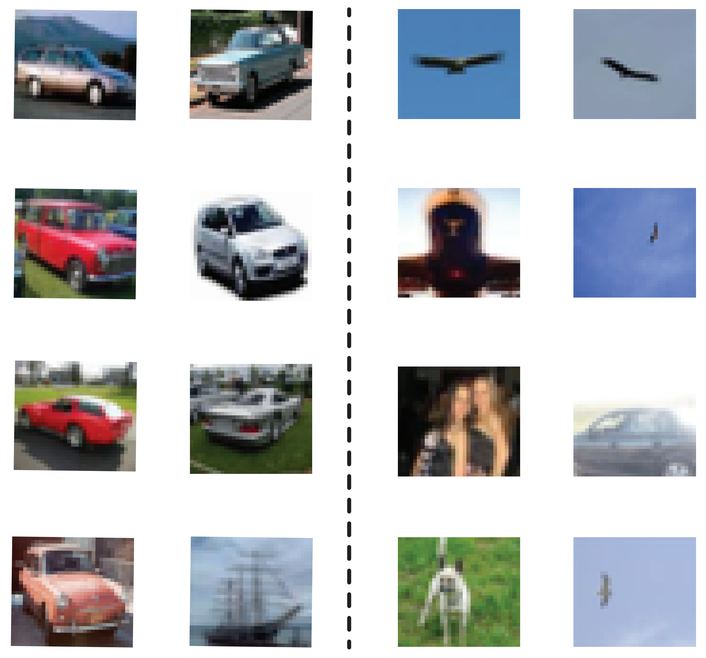
- Baseline does not filter examples
- SB filters >55% of CIFAR10 examples
- SB reaches target errors with fewer iterations
- SB improves final test error

SVHN (w/ Label Error)

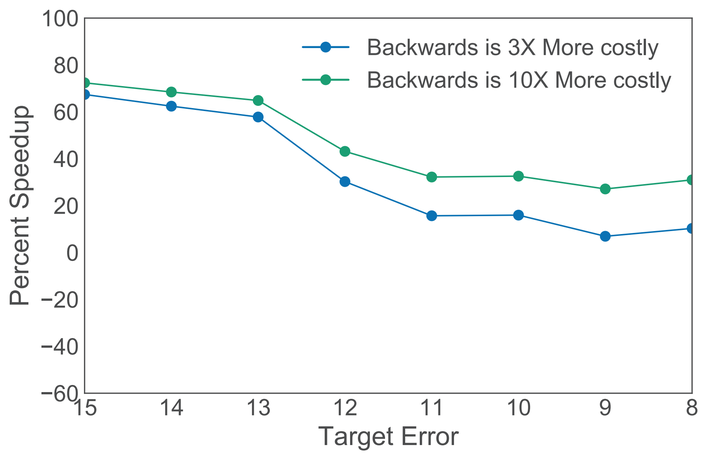


- SB filters > 80% of SVHN examples
- SB reaches target errors with fewer iterations
- SVHN is known to have label error
- SB reaches same final accuracy despite label error

Diving into CIFAR10

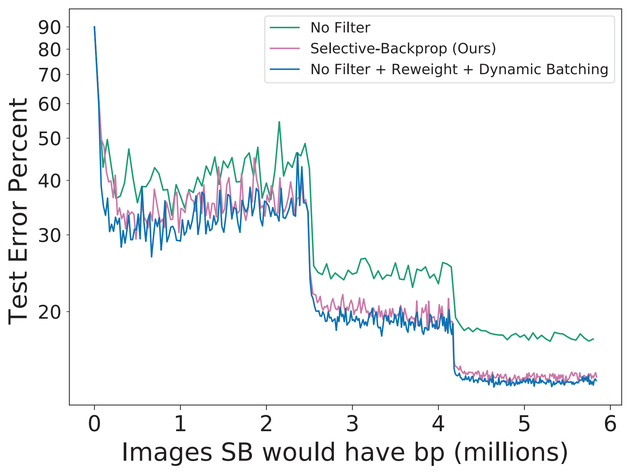


Cifar10 Speedup

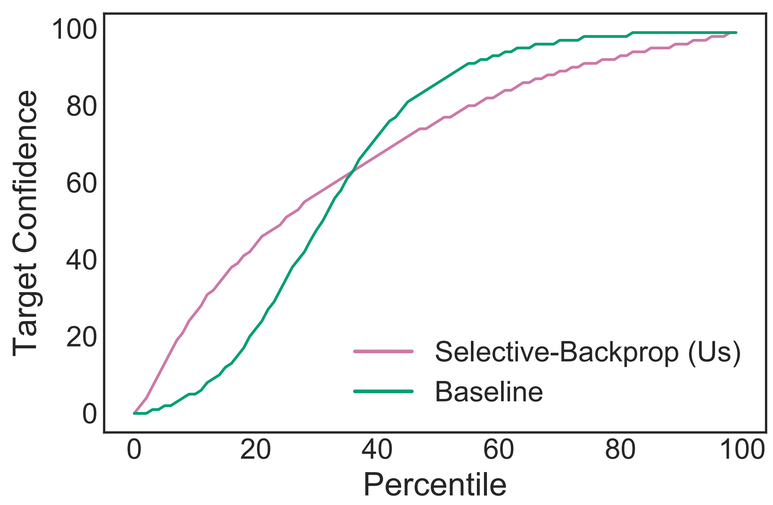


- Relative speedup compared to baseline dependent on:
  - Relative cost of backwards and forwards pass
  - Target error
- SB gives up to 78% speedup (theoretical max is 100%)

Selective-Backprop Benefit Attribution



- SB is equivalent to NoFilter with:
  - Larger batch sizes
  - Downweighting losses based on prob



- Y-axis is confidence in our predicition of correct class
  - On test examples during snapshot of training
- SB improves confidence of challenging examples
  - Without sacrificing confidence of easier examples