Faster DNN Training With Selective Backpropagation

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

Background

Can we speed up training by only backpropogating useful examples?

Motivation

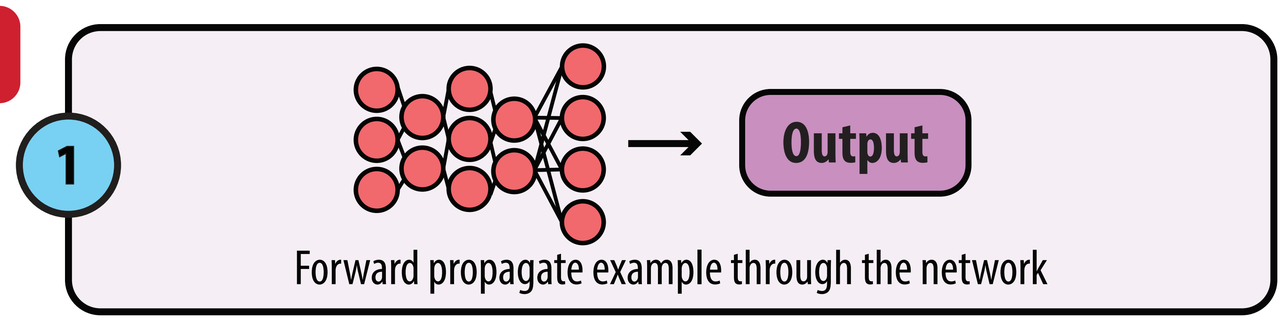
- Train network to target accuracy quickly
- Get a faster signal from training (e.g., for hyperparameter search)
- Cherry-pick useful training examples from ever-growing datasets

Opportunity

- Fast inference (e.g., with accelerators) => explore dataset cheaply

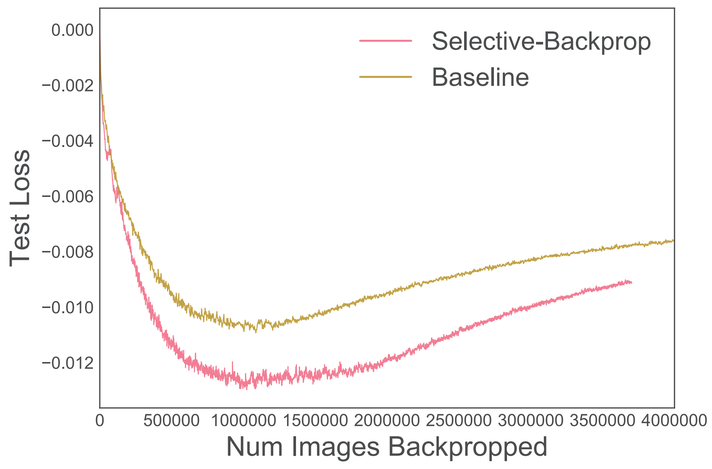
Approach

- Isolate useful examples using inference (the output of the forward pass)
- Reduce cost of inference using hardware-accelerators or model compression
- Reduce no. of backwards passes by only training on "surpising" examples

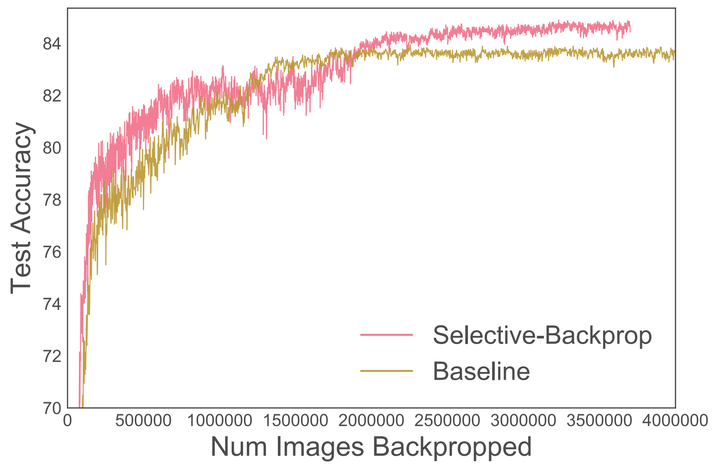


Selective Backprop gives more accuracy with fewer training examples

CIFAR10

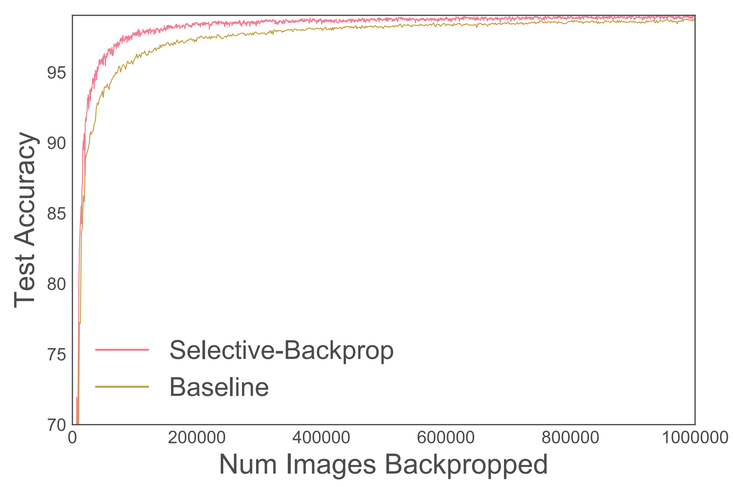


- Baseline does not filter examples
- Selective Backprop (SB) filters >45% of examples



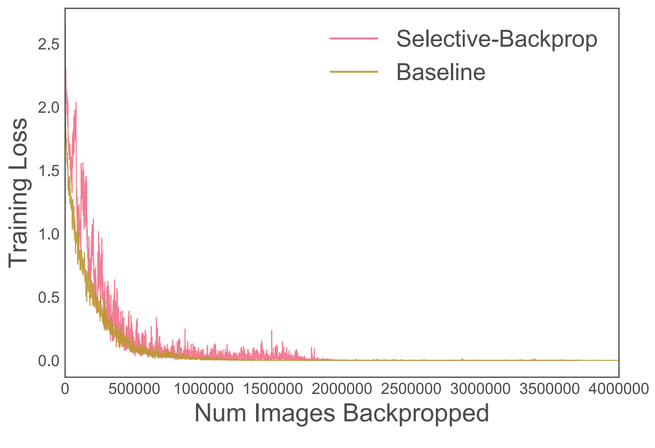
- SB reduces test loss with fewer examples
- Achieves X% of accuracy with Y fewer examples

MNIST



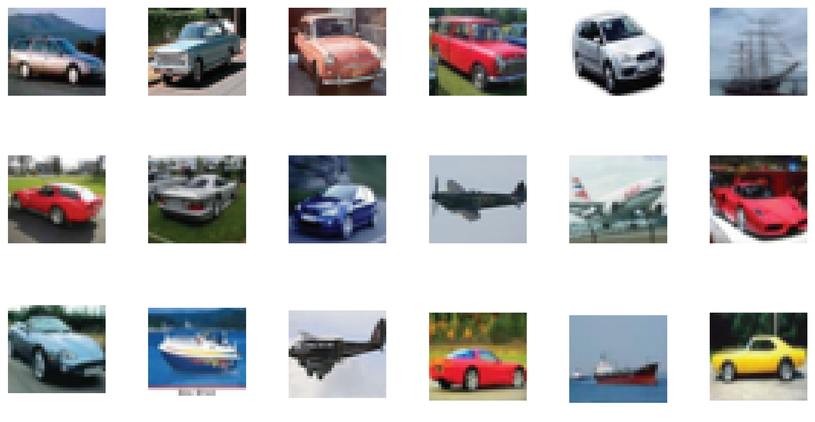
- SB filters 85% of examples of MNIST
- Achieves X% of accuracy with Y fewer examples

Diving into CIFAR10

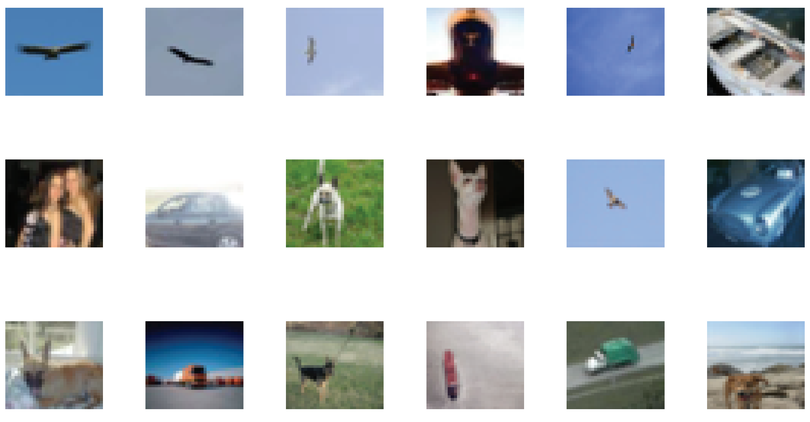


SB chooses to train images with higher loss

Easy Examples



Hard Examples



Next Steps: Use Selective Backprop to Improve Wall-Clock Time