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

Goal: Improve wall-clock time of training to a particular accuracy

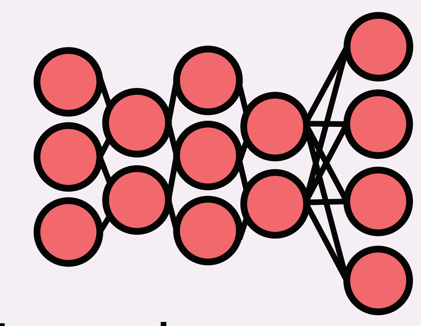
Trends
- Labeled datasets are larger (e.g., high frame rate video, click through data)
- DNN inference is faster with accelerators and deep compression

Approach
- Isolate hard examples using output of the forward pass
- Reduce cost of backwards pass by only training on hard examples
- Reduce cost of inference using hardware-accelerated or quantized inference

Motivation
- Choose learning examples more efficiently (e.g., to filter redundant datasets)
- Get signal from training quickly (e.g., hyperparameter search)

Can we speed up training by only backpropogating "surprising" examples?

1



→ Output

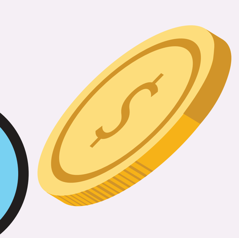
Forward propagate input through the network

2

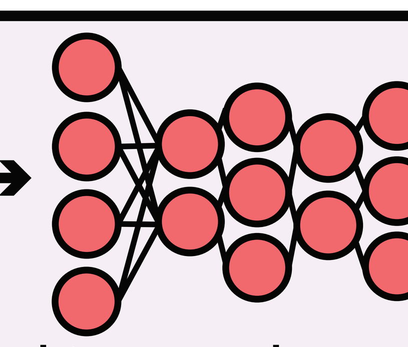
P(Backprop) = L2 Dist²(Output , Target)

Calculate

3



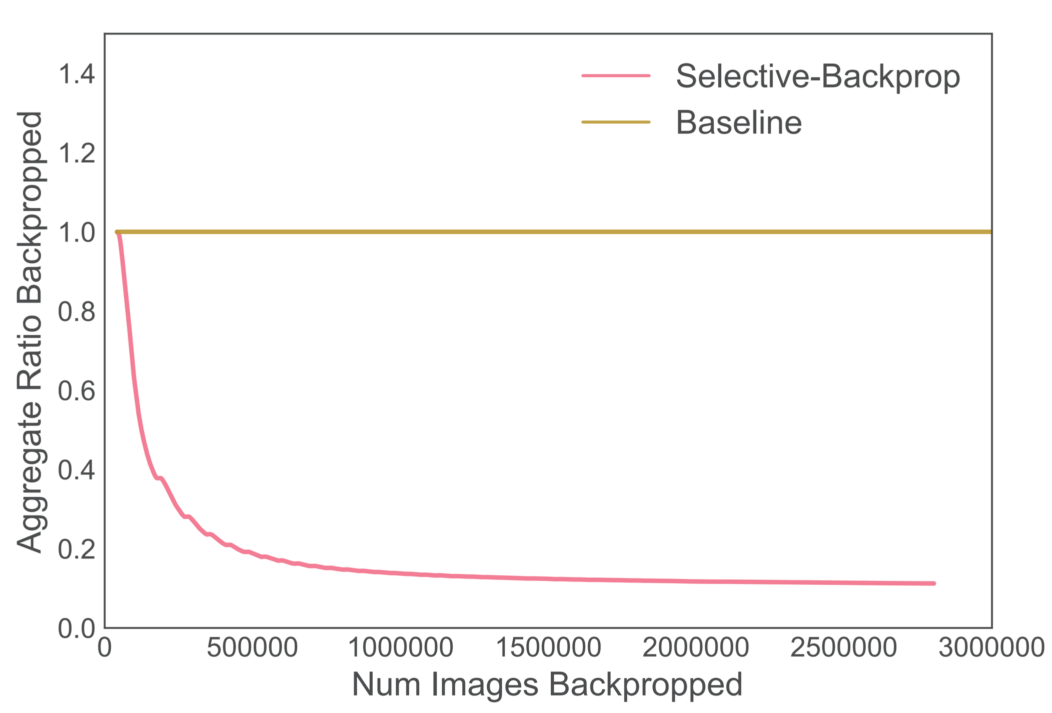
→ Output → Loss →

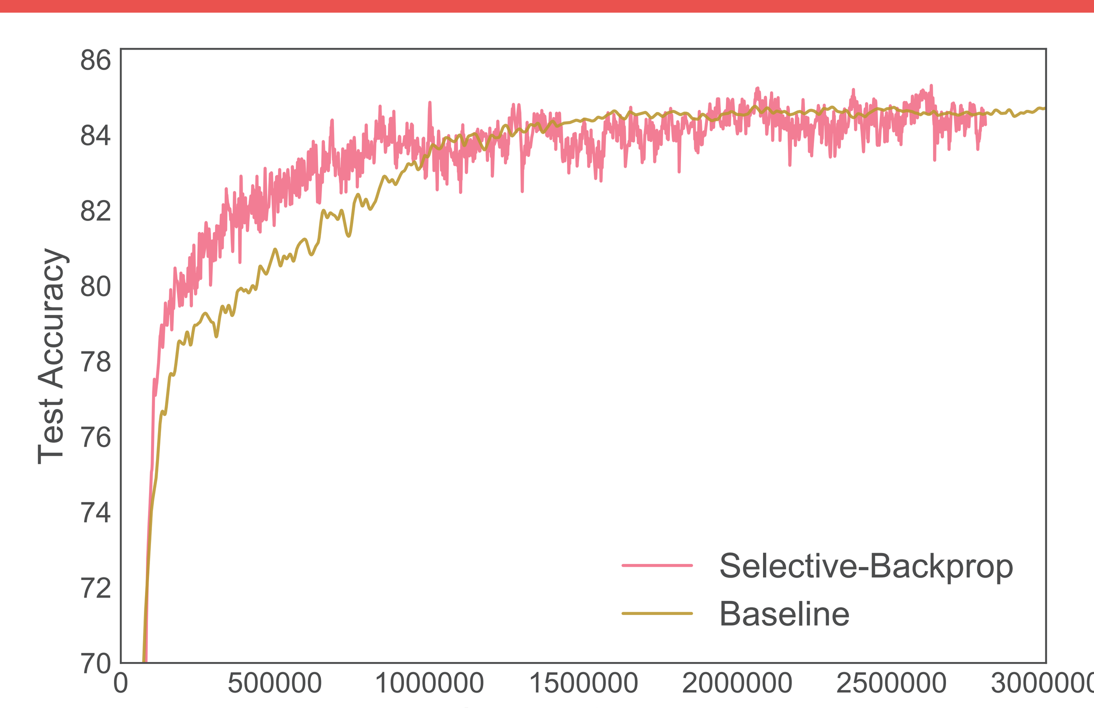


Flip a coin to determine if we should backprop this example

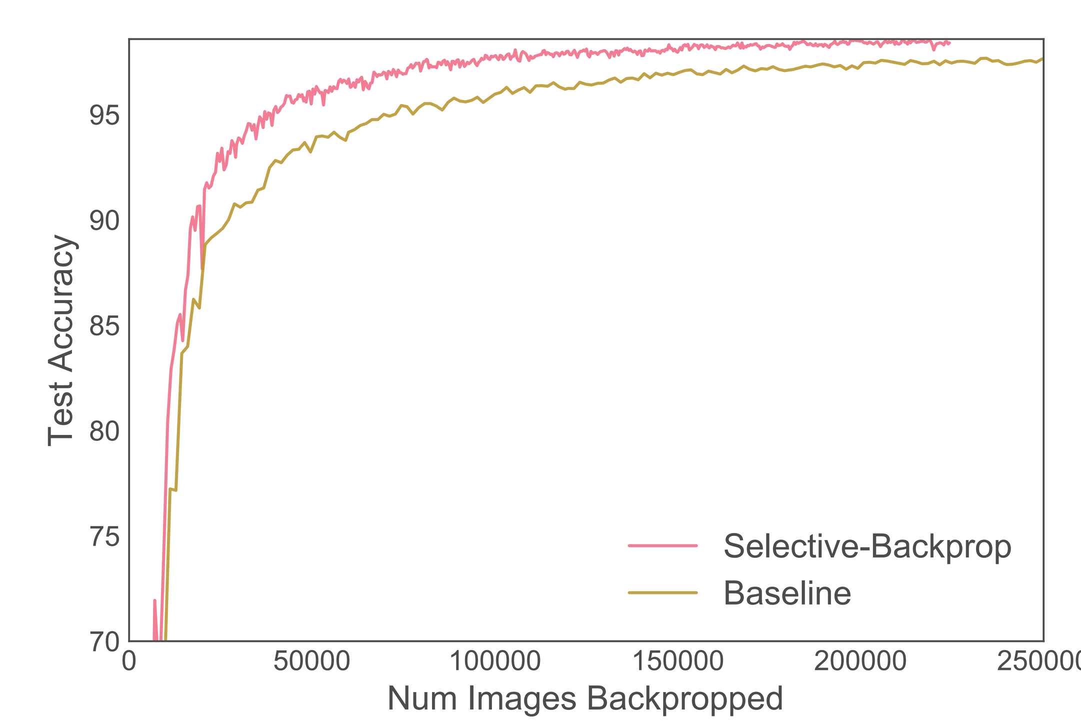
Selective Backprop achieves same accuracy with fewer training examples

CIFAR10





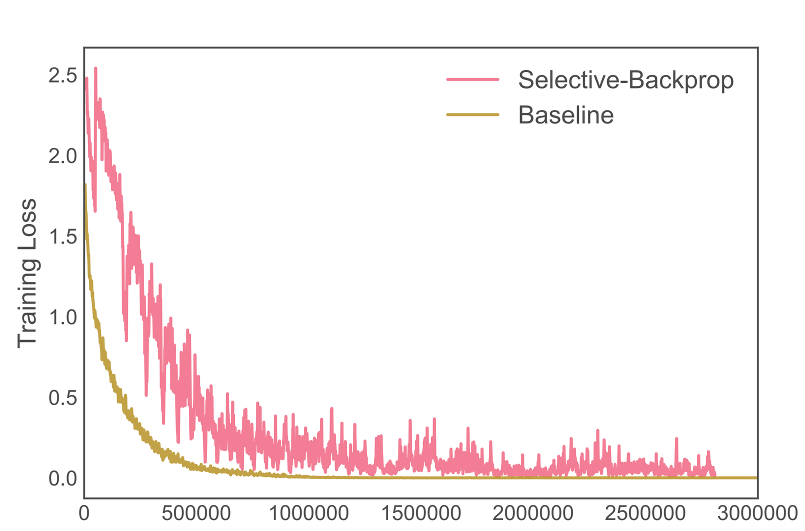
MNIST

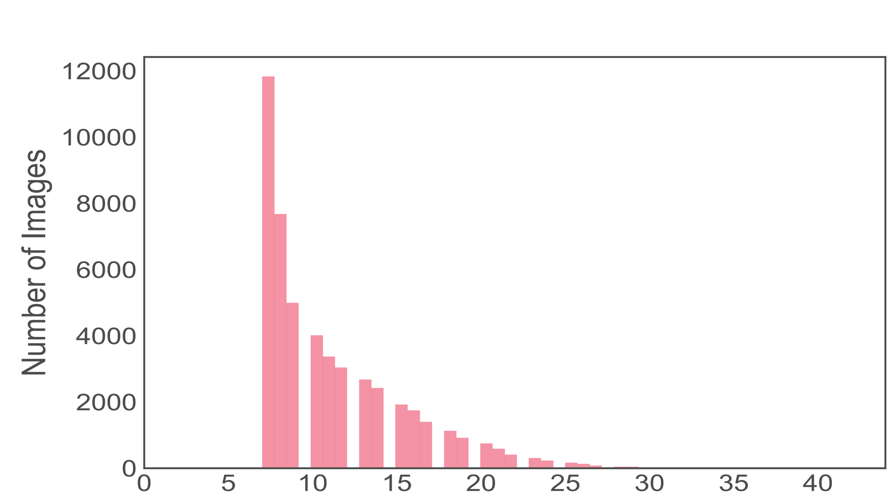


- Baseline does not perform filtering
- Selective Backprop (SB) filters >85% of examples

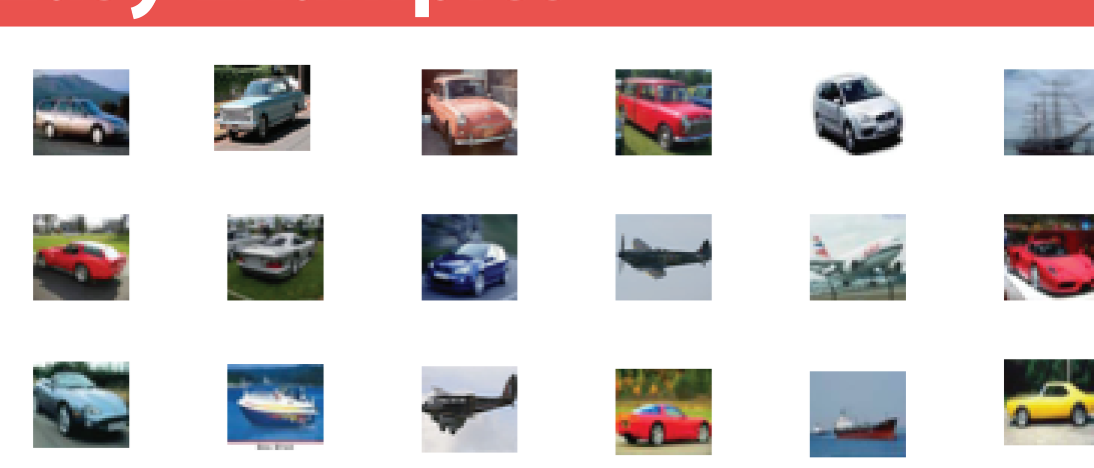
- SB achieves comparable accuracy with fewer examples on both CIFAR10 and MNIST
- SB reduces test loss quicker than baseline

Diving into CIFAR 10

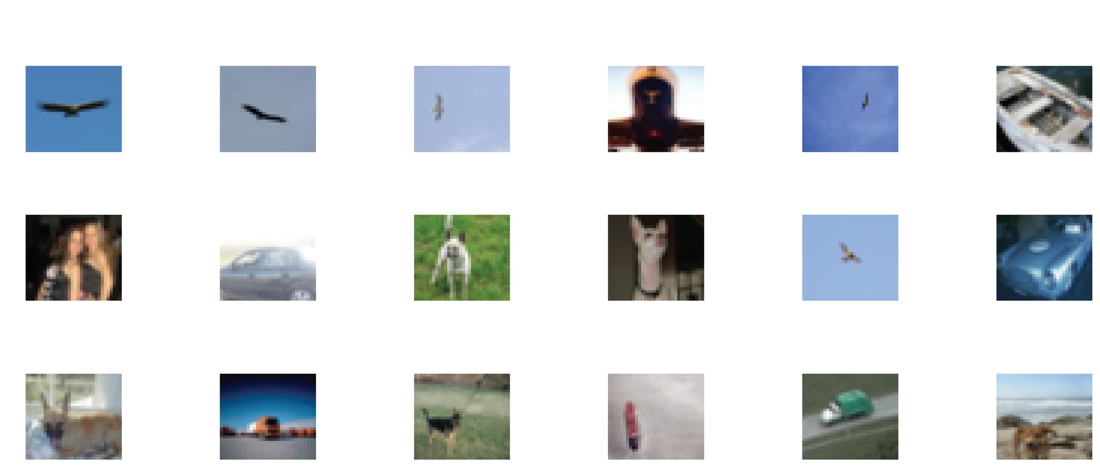




Easy Examples



Hard Examples



Future Work