Improving Image Classifiers for Small Datasets by Learning Rate Adaptations

Sourav Mishra[1] Toshihiko Yamasaki[1] Hideaki Imaizumi[2]

東京大学 THE UNIVERSITY OF TOKYO

[1]*The University of Tokyo, Bunkyo-ku, Tokyo*
[2]*exMedio Inc., Chiyoda-ku, Tokyo*

exMedio

OVERVIEW

Issues with SOTA image classification

- Multiple iteration for SOTA
- Longer training times
- Higher billed costs

Proposed Solution

- Combining proven techniques on standard datasets.
- Optimization by Learning Rate (α) adaptation.
- Transfer modus-operandi to custom applications.

LEARNING BASELINE

- CIFAR-10 Multi-class classification mimicking custom task.
- Choice of 6 architectures of increasing size & density.

  ResNet-34, ResNet-50, ResNet-101

  ResNet-152, DenseNet-161
- Early stopping & manual α selection (0.01, 0.001)
- Training up to at least 90% of accuracy.

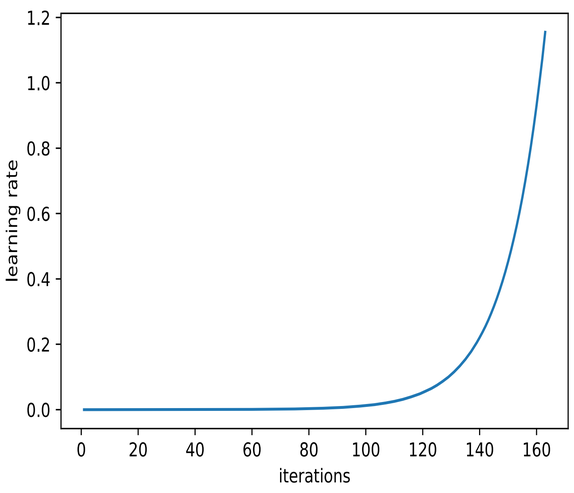
Architecture	Accuracy (Top-1)	Time (s)
ResNet 34	90.36%	17,757
ResNet-50	90.54%	34,039
ResNet-101	90.71%	60,639
ResNet-152	90.68%	91,888
DenseNet-161	93.02%	54,628

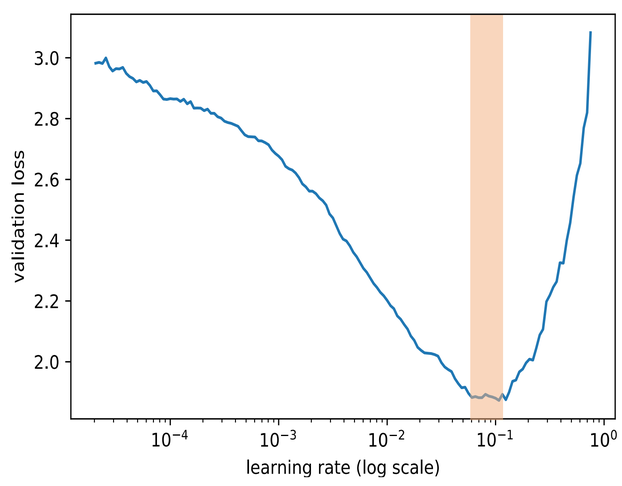
PROPOSAL

LEARNING RATE (LR) RANGE TEST

Using several minibatches, increase the Learning rate gradually.
Observe the cost function vis-à-vis LR:

- The cost function dips slowly at small LR
- Increases dramatically in a small region.
- Reverses direction beyond certain value.

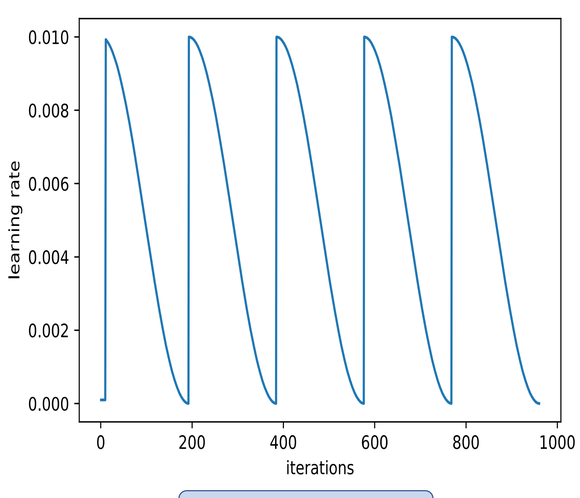




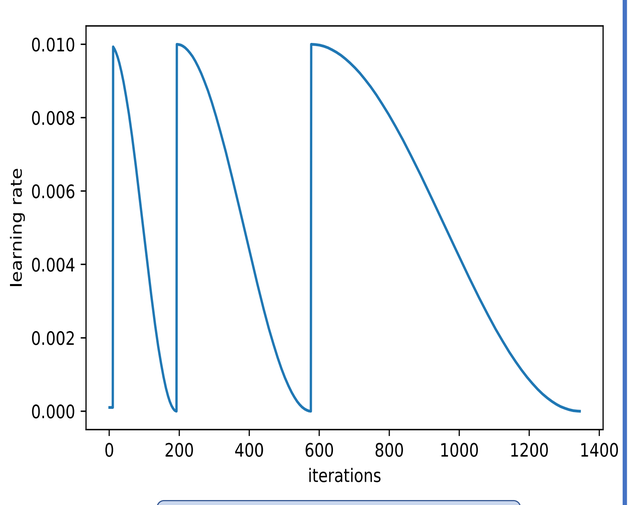
SGD-R WITH COSINE ANNEAL & DERIVATES

Stage 1: SGD-R + Cosine rate cycle (FC Layer)

Stage 2: SGD-R + Cosine cycle + Layer-dependent rate (Full Net)

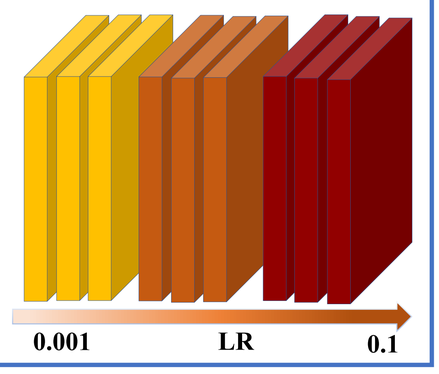


SGD-R



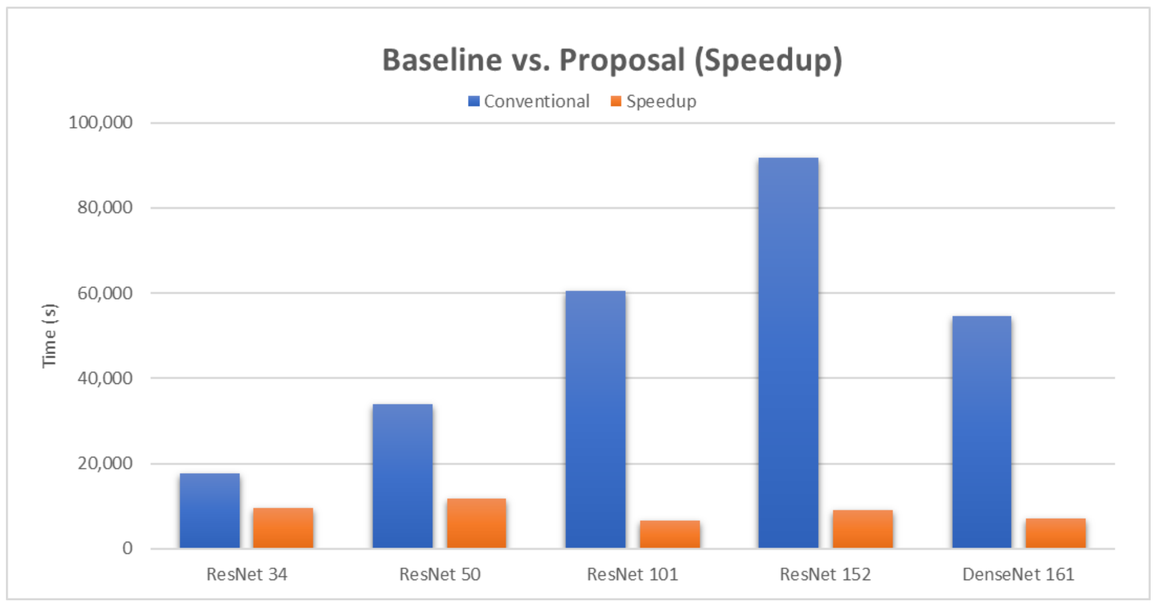
SGD-R + CLM + DLR

- SGD-R with Cosine rate gives a good fit quickly.
- SGD-R + Cycle length multiplication & Differential learning rates fits the model tightly with minimal disturbances.
- Number of epochs for entire process is exponentially lower than conventional training.

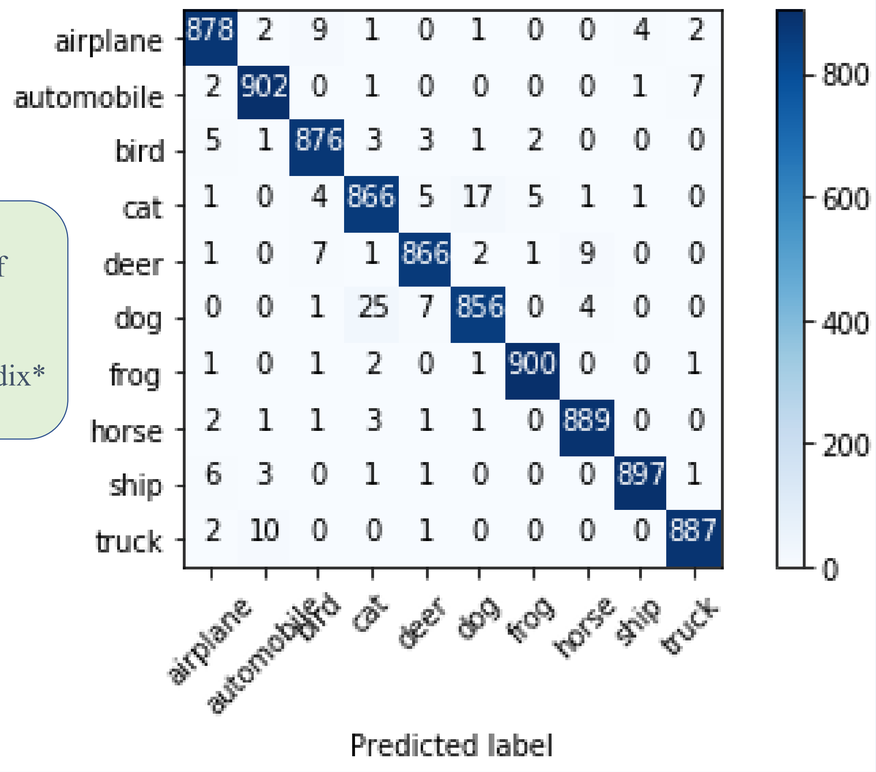


EXPERIMENTAL RESULT

Architecture	Accuracy (Top-1)	Time (s)
ResNet 34	96.84%	9,565
ResNet-50	96.82%	11,817
ResNet-101	97.61%	6,673
ResNet-152	97.78%	9,012
DenseNet-161	97.15%	7,195



- Confusion matrix of **ResNet-152**
- Available in Appendix*



APPLICATION

Dermatological classification

- Homogenous, small dataset
- Continuous re-train required.

10%+ accuracy gain
3.1X to 5.7X speedup

$ 6.89 per train (conventional)
vs.
$0.70 per train (speedup method)

** 2019 AWS EMR p2.xlarge*