Supplementary Material: Learning Rate Optimization for Online Deep Learning

Anonymous[1]

No Institute Given

Hyperparameter Settings for Empirical Evaluation

Table 1: Parameter Notation as used in Paper.

Parameter	Symbol
Learning Rate	η
Decay Factor	γ
Drift Detection Confidence Level	δ
Steps Between LR Cycles/Updates	s
Relative LR at Midpoint of Cycle	$\hat{\eta}$

Table 2: Search spaces for learning rate.

Optimizer	Learning Rate Search Space
SGD	$2^1, 2^0, ..., 2^{-8}$
Adam	$2^{-3}, 2^{-4}, ..., 2^{-12}$
AdaGrad	$2^1, 2^0, ..., 2^{-8}$
WNGrad	$10^{1.25}, 10^{0.75}, ..., 10^{-7.75}$
HD	$2^{-3}, 2^{-4}, ..., 2^{-12}$
COCOB	100
DoG	1
D-Adapt	1
Mechanic	0.01

Table 3: Other parameter settings.

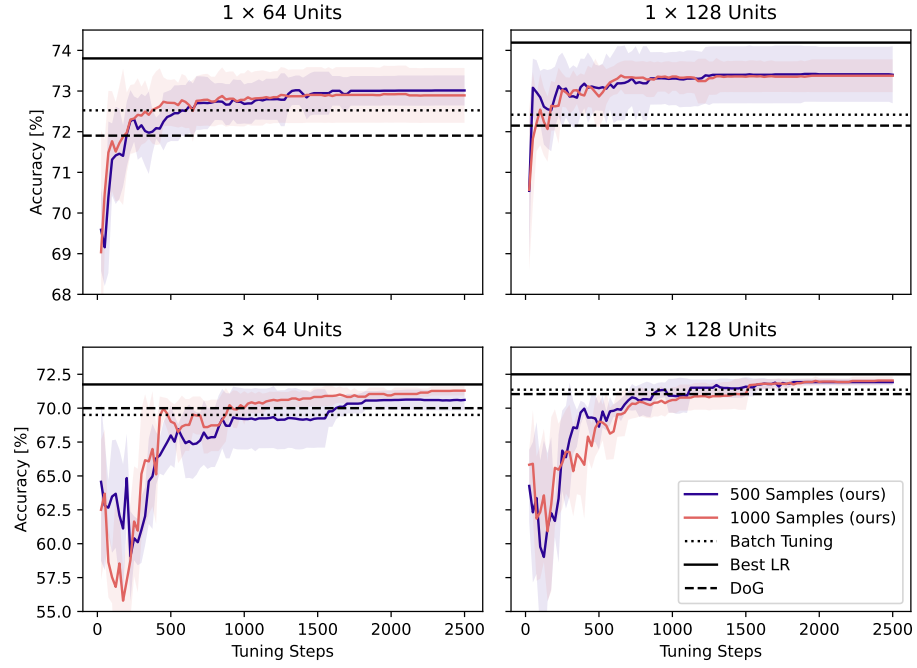| Schedule	Values
Exponential	$\gamma = 1 - 2^{-13}$
Exp. Reset	$\gamma = 1 - 2^{-12}, \delta = 0.0001$
Step	$\gamma = 0.75, s = 2000$
Cyclic	$\hat{\eta} = 0.25, s = 8000$

Additional Results for Pre-Tuning

Fig. 1: Average accuracy over all evalutated real-world datasets achieved by pre-tuning for different network architectures.