“To ensure a fair comparison, we tune the peak learning rate lr and decoupled weight decay λ for both AdamW (Adafactor) and our Lion using a logarithmic scale” in the LION paper

1. **SGD**

hyperparams = {

“lr” : [1e-2, 5e-3, ],

“weight\_decay” : [e-2, e-3],

“momentum” : [],

}

https://arxiv.org/pdf/1907.08931v2.pdf

https://lightning.ai/docs/pytorch/stable/notebooks/lightning\_examples/cifar10-baseline.html

<https://nni.readthedocs.io/en/v2.0/TrialExample/Cifar10Examples.html>

<https://wandb.ai/gautamsagar/pytorch-assignment/reports/Training-ResNet18-on-CIFAR10---VmlldzoxMDA5MzAx>

<https://sgugger.github.io/the-1cycle-policy.html>

<https://medium.com/mlearning-ai/designing-convolutional-neural-networks-on-cifar10-dataset-using-transfer-learning-resnet101-423473d1b17c>

<https://github.com/rasbt/deeplearning-models/blob/master/pytorch_ipynb/cnn/cnn-resnet101-cifar10.ipynb>

https://towardsdatascience.com/the-best-learning-rate-schedules-6b7b9fb72565

https://openaccess.thecvf.com/content/CVPR2022/supplemental/Hosseini\_Exploiting\_Explainable\_Metrics\_CVPR\_2022\_supplemental.pdf

LR search in [0.1,0.05, 0.01,0.005, 0.001,0.0005, 0.0001] resnet18

weight decay [e-2,e-3,e-4,e-5] resnet 18 et resnet 56

lr in [0.1, 00.1, 0.000057] resnet101 :/

1. **SGD with Momentum**

https://arxiv.org/pdf/2103.17182.pdf

Lr 0.0001, 0.001, 0.01, 0.1, 1, 10

momentum [0.5,0.6,0.7,0.8,0.9, 0.95] resnet18

wd studied<e-4, optimal= 0.0005

1. **RMS for cifar10**

<https://csyhquan.github.io/manuscript/19-pr-Barzilai%E2%80%93Borwein-based%20adaptive%20learning%20rate%20for%20deep%20learning.pdf>

lr in [0.01,e-3,e-4, e-5, e-6]

wd in e-4, 5e-4

1. **ADAM**

“To ensure a fair comparison, we tune the peak learning rate lr and decoupled weight decay λ for both AdamW (Adafactor) and our Lion using a logarithmic scale” in the LION paper

Lr, WeightDecay : 3e− 3, 0.1 when trained on ImageNet

1. **LION**

“To ensure a fair comparison, we tune the peak learning rate lr and decoupled weight decay λ for both AdamW (Adafactor) and our Lion using a logarithmic scale” in the LION paper

Lr, WeightDecay: 3e 4, 1.0 when trained on ImageNet

**Scheduler**

“By default, the learning rate schedule is cosine decay with 10K steps warmup” in the LION paper.

1. **SGD cifar100**

<http://people.ee.duke.edu/~lcarin/6646_supp.pdf>

https://jovian.com/rensortino/cifar100-resnet18

<https://towardsdatascience.com/the-best-learning-rate-schedules-6b7b9fb72565>

https://www.epfl.ch/labs/mlo/wp-content/uploads/2021/05/crpmlcourse-paper841.pdf

LR search in [0.1,0.05, 0.01,0.005, 0.001,0.0005, 0.0001] resnet18

weight decay [e-2,5e-3,5e-5] resnet 18

lr in [0.1, 00.1, 0.000057] resnet101 :/

1. **SGD with Momentum**

<https://towardsdatascience.com/why-0-9-towards-better-momentum-strategies-in-deep-learning-827408503650>

Lr =e-1,e-2,e-t3,e-4

WeightDecay pretty much the same as before but with momentum [0.5,0.6,0.7,0.8,0.9, 0.95] resnet18

8)**RMS cifar100**

[**https://www.osti.gov/servlets/purl/1671416**](https://www.osti.gov/servlets/purl/1671416)

**lr=0.0001**

**wd=e-6**

**momentum=0.9**