Training ResNet8 in less than 10 minutes:

We used the following optimizations:

1. Preprocessing,Data Augmentation,Batching on GPU
2. Maxpool before ReLU & Batchnorm layers to reduce computations in these two layers
3. Label smoothing to create uniform distribution of target
4. CELU activation instead of ReLU layer to improve generalization
5. Input patch whitening to control covariance of internal layers
6. Exponential moving averages
7. Test-time augmentation
8. Using SAM optimizer instead of SGD

Hyperparameters:

- Lr = 0.1
- Batch size = 256
- Label smoothing = 0.1
- Weight decay = 0.0005

- Optimizer = SAM

- lr scheduler = PiecewiseLinear