Optimizer Wall Clock Time vs. Model Size (CUDA)

Average Step Time (ms) vs. Number of Model Parameters

Optimizer
- Amsgrad
- Adagrad
- Adam
- Adamp
- Adan
- Diffgrad
- Lion
- Madgrad
- Pulsegrad
- Rmsprop