

Normalized Training Loss

10^{-1}

10^{-2}

- 8x A100
- 32x A100
- CS1 (mixed precision)

0 20 40 60 80 100 120

Cumulative Training Time / min

