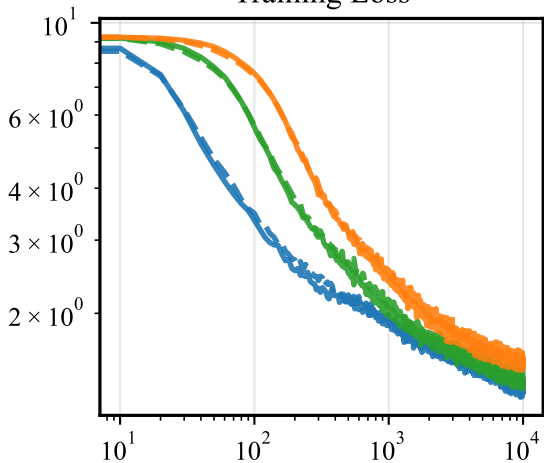


Training Loss



Training Steps

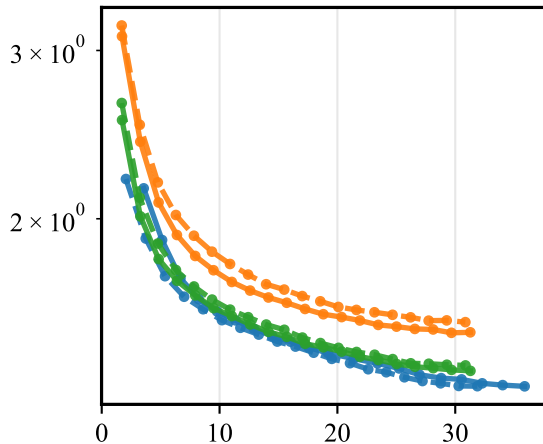
— lr = 0.01 SwiGLU

- - lr = 0.01 SiLU

— lr = 0.001 SwiGLU

- - lr = 0.001 SiLU

Validation Loss



Relative Time (min)

— lr = 0.0003 SwiGLU

- - lr = 0.0003 SiLU