

Loss of 1L Transformer on Modular Addition ($p=97$)
 $lr=1e-1$, $weight_decay=0.001$

