validation loss evolution @ add benchmark 10⁰ memory augmented transformer 5×10^{-1} differentiable neural computer unitary rnn matrix exponential unitary rnn 2×10^{-1} transformer recurrent_network_attention_transformer 10^{-1} recurrent_network_augmented_transformer gru $\begin{array}{c} 5 \times 10^{-2} \\ p \\ p \\ s \\ u \\ s \\ u \\ s \\ 0 \\ 2 \times 10^{-2} \\ 10^{-2} \\ 5 \times 10^{-3} \\ 2 \times 10^{-3} \end{array}$ 5×10^{-2} neural circuit policies ct rnn ct gru ode Istm unitary_ncp 10^{-3} 5×10^{-4} 2×10^{-4} 10^{-4} 0 6 12 18 24 30 36 42 48 54 epochs