

model	trainable parameters	training duration total [s]	training duration per epoch [s]	epochs	test mean squared error
transformer	21889	421.042 ± 62.125	33.175 ± 3.885	12.667 ± 0.577	0.000 ± 0.000
recurrent_network_augmented_transformer	3729	3369.947 ± 238.022	198.209 ± 6.384	17.000 ± 1.000	0.000 ± 0.000
gru	20241	423.332 ± 52.080	20.821 ± 0.322	20.333 ± 2.517	0.001 ± 0.000
ct_gru	19073	983.703 ± 123.086	48.464 ± 0.901	20.333 ± 2.887	0.001 ± 0.000
recurrent_network_attention_transformer	2349	10433.169 ± 3205.297	370.905 ± 13.507	28.000 ± 7.937	0.002 ± 0.002
differentiable_neural_computer	24774	6138.672 ± 4012.017	152.861 ± 2.424	40.333 ± 26.858	0.007 ± 0.011
ct_rnn	17025	9904.616 ± 1578.008	193.094 ± 1.600	51.333 ± 8.505	0.019 ± 0.003
matrix_exponential_unitary_rnn	17409	5956.370 ± 1220.646	115.400 ± 11.578	51.333 ± 6.028	0.022 ± 0.032
lstm	17217	589.247 ± 181.783	19.420 ± 0.162	30.333 ± 9.292	0.066 ± 0.085
unitary_rnn	2929	9694.727 ± 2483.450	278.442 ± 11.350	35.000 ± 9.849	0.094 ± 0.076
unitary_ncp	1885	3283.152 ± 1513.217	102.520 ± 0.625	32.000 ± 14.731	0.106 ± 0.092
memory_augmented_transformer	18066	3802.347 ± 1303.816	283.514 ± 8.116	13.333 ± 4.163	0.122 ± 0.083
neural_circuit_policies	1349	1007.356 ± 252.374	125.919 ± 0.414	8.000 ± 2.000	0.166 ± 0.001
ode_lstm	25537	748.046 ± 108.977	44.046 ± 0.415	17.000 ± 2.646	0.167 ± 0.002