recurrent network augmented transformer @ MnistBenchmark 2.2 0.7 2.0 1.8 0.6 1.6 rain sparse categorical crossentropy train sparse categorical accuracy 0.5 val sparse categorical crossentropy val sparse categorical accuracy 1.4 test sparse categorical crossentropy test sparse categorical accuracy 0.4 1.2 1.0 0.3 0.8 0.2 8 16 56 0 24 32 40 48 64 72 epochs