recurrent network augmented transformer @ ActivityBenchmark 0.9 1.6 0.8 1.4 0.7 1.2 train spars categorical crossentropy train sparse categorical accuracy 1.0 val sparse categorical crossentropy val sparse categorical accuracy test spare categorical crossentropy test sparse categorical accuracy 0.6 8.0 0.5 0.6 0.4 0.4 5 0 10 15 20 25 30 35 epochs