recurrent network attention transformer @ memory benchmark 1.0 1.4 1.2 0.9 1.0 0.8 0.8 train sparse categorical crossentropy train sparse categorical accuracy val sparse categorical crossentropy val sparse categorical accuracy test sparse categorical crossentropy test sparse categorical accuracy 0.6 0.7 0.4 0.6 0.2 0.0 0.5 3 6 9 12 15 18 21 24 27 epochs