recurrent network augmented transformer @ memory benchmark 1.0 2.0 0.9 0.8 1.5 0.7 train sparse categorical crossentropy train sparse categorical accuracy val sparse categorical crossentropy val sparse categorical accuracy F 0.6 1.0 test sparse categorical crossentropy test sparse categorical accuracy 0.5 0.4 0.5 0.3 0.0 0.2 3 6 9 12 15 18 21 24 27 epochs