recurrent network attention transformer @ mnist benchmark 1.4 0.85 sparse categorical crossentropy 0.80 1.2 sparse categorical accuracy 0.75 1.0 train sparse categorical crossentropy train sparse categorical accuracy val sparse categorical crossentropy val sparse categorical accuracy 0.70 test sparse categorical crossentropy test sparse categorical accuracy 8.0 0.65 0.60 0.6 0.55 0.4 3 6 9 12 15 18 21 24

epochs