

Model	Size	Training	Val. perp.	Test perp.
Vanilla RNN	20M parameters		103.0	97.7
GRU	20M parameters		95.5	91.7
LSTM	20M parameters		83.3	78.8
LSTM (2 layers)	20M parameters	Trained by us	78.4	74.3
CFN (2 layers)	20M parameters	Trained by us	79.7	74.9
LSTM (2 layers)	50M parameters	Trained by us	75.9	71.8
CFN (2 layers)	50M parameters	Trained by us	77.0	72.2