Gru Lstm DecayRnn (ON Lstm) on difficult tasks as well

CNN results

see why decay rnn grads blow up in Predict VerbNumber tasks

Have a look at gate parameters - **Hadamard prods to scalars**