Visualizing the behavior of LSTM and GRU when performing speech recognition tasks:
1. Activation patterns
2. Temporal trace
3. Memory robustness

Modifications inspired by the visualization:
1. lazy cell update in LSTM
2. shortcut connections for residual learning

LSTM & GRU

Visualization

Modifications

Experiment Results

Conclusions

LSTM and GRU use different ways to encode information and the information in GRU is more distributed. LSTM possesses a long-term memory but it is also noise-sensitive.

Inspired by these observations, we introduced two modifications to enhance gated RNNs: lazy cell update and short connections for residual learning, and both provide interesting performance improvement.