

IDLAB, GHENT UNIVERSITY - IMEC

Frédéric Godin, Joni Dambre and Wesley De Neve

IMPROVING LANGUAGE MODELING USING DENSELY CONNECTED RECURRENT NEURAL NETWORKS

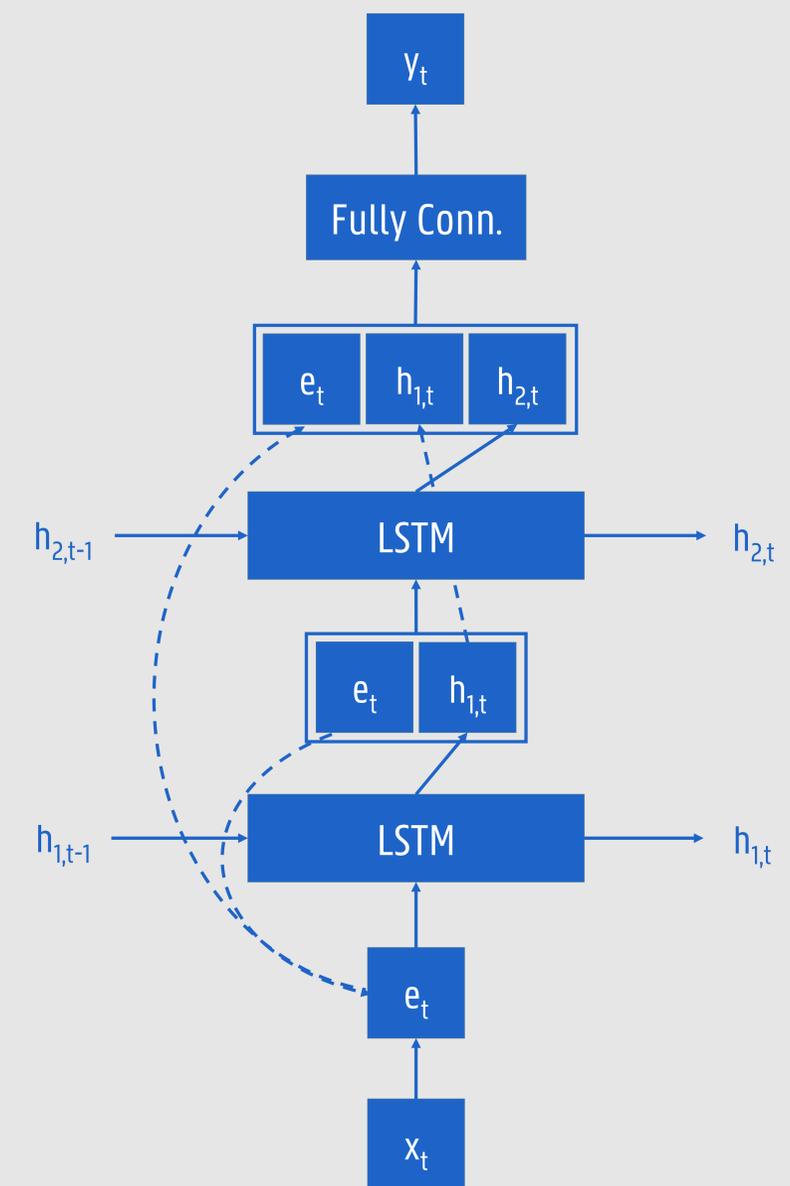
MOTIVATION

- Skip or residual connections are only sporadically used when stacking LSTMs
- Densely connecting all layers with skip connections is very successful in convolutional neural networks

RESEARCH QUESTION

- What if we add a skip connection between every output and every input of every layer in a recurrent neural network?

ARCHITECTURE



EXPERIMENTS

Model	Hidden states	# Layers	# Params	Perplexity
Stacked LSTM (Zaremba et al., 2014)	650	2	20M	82.7
	1500	2	66M	78.4
Stacked LSTM	200	3	5M	108.8
	350	2	9M	87.9
Densely Connected LSTM	200	2	9M	80.4
	200	3	11M	78.5
	200	4	14M	76.9

Evaluated on the Penn Treebank word-level language modeling task. Test set perplexity is reported.

CONCLUSION

- ➔ Densely connecting all layers substantially improves language modeling performance
- ➔ We use six times fewer parameters to obtain the same result as a stacked LSTM