**time-frequency (TF) LSTM (2D):**

[35] J. Li, A. Mohamed, G. Zweig, and Y. Gong, “LSTM time and frequencyrecurrence for automatic speech recognition,” inProc. ASRU, 2015.

[36] ——, “Exploring multidimensional LSTMs for large vocabulary ASR,”inProc. ICASSP, 2016.

[37] T. N. Sainath and B. Li, “Modeling time-frequency patterns withLSTM vs. convolutional architectures for LVCSR tasks,” inProc.INTERSPEECH, 2016.

Although bi-directional LSTMs (BLSTMs) perform betterthan uni-directional LSTMs by using the past and futurecontext information [8], [42], they are not suitable for real-timesystems since the recognition can happen only after the wholeutterance has been observed. For this reason, models, suchas latency-controlled BLSTM (**LC-BLSTM**) [29] and row-convolution BLSTM (**RC-BLSTM**), that bridge between uni-directional LSTMs and BLSTMs have been proposed. In thesemodels, the forward LSTM is still kept as is. However, thebackward LSTM is replaced by either a backward LSTM withat mostN-frames of lookahead as in the LC-BLSTM case,or a row-convolution operation that integrates information intheN-frames of lookahead. By carefully choosingNwe canbalance between recognition accuracy and latency. Recently,LC-BLSTM was improved by [43] to speed up the evaluationand to enable real-time online speech recognition by usingbetter network topology to initialize the BLSTM memory cellstates.

http://www.jmlr.org/papers/volume3/gers02a/gers02a.pdf

**Peephole connections**.Our simple but effective remedy is to add weighted “peephole”connections from the CEC to the gates of the same memory block (Figure 2). The peepholeconnections allow all gates to inspect the current cell state even when the output gate isclosed. This information can be essential for finding well-working network solutions ,as wewill see in the experiments below.