- RNNs (Recurrent Neural Networks) add loops to allow persistent information
- Used for speech recognition, language modeling, translation, image captioning, …
- Works best when the distance to the relevant context information is small
  - In theory our Physics Engine predictions should only need to look at the last two frames at most which should be a small enough gap to learn from
- LSTMs (Long Short Term Memory networks) can learn context clues over a much larger distance.
- LSTMs have a cell state memory that persists through the repetitions of the RNN module
- This cell state memory can be changed, dependent on gates
- Gates consist of a sigmoid neural network layer and a pointwise multiplication that outputs values between 0 and 1 that describe how much of each component of the cell state should pass to the next repetition
- LSTMs have three gates
- The first gate takes the last repetitions output and the new input and calculates how much of the cell state memory should be kept
  - E.g Objects that are in motion should have their old positions forgotten
- The second gate calculates which values should be updated and what they should be updated to and adds them to the cell state
  - E.g objects in motion should have their new positions calculated and added
- The third gate looks at the updated cell state and decides what the final output should be
- Variants on LSTMs exist that change what information each gate has access to and each performs differently on different tasks
  - The Gated Recurrent Unit is simpler than a standard LSTM and may be useful for the initial stages of implementation. Specialised models should be looked into if improvement stalls but shouldn't be needed until the complexity of the simulation increases.

http://colah.github.io/posts/2015-08-Understanding-LSTMs/