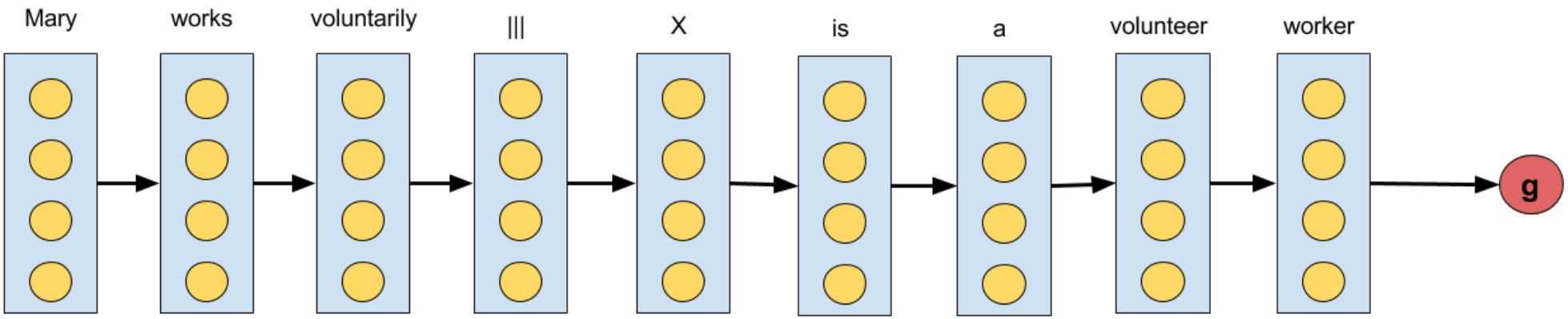
Gated RNNs

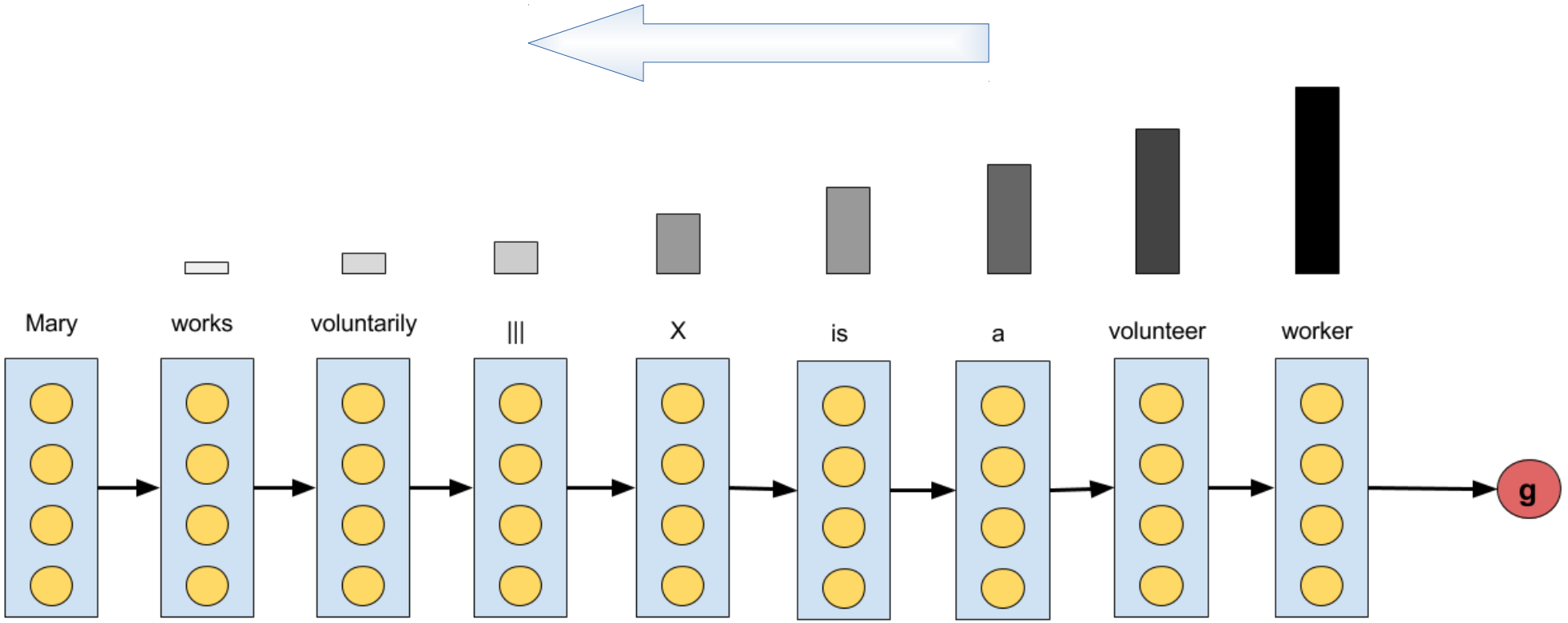
A Battle Against Vanishing Gradients

RNN

$$\vec{h}_{t+1} = tanh(W\vec{x}_{t+1} + U\vec{h}_t + b)$$



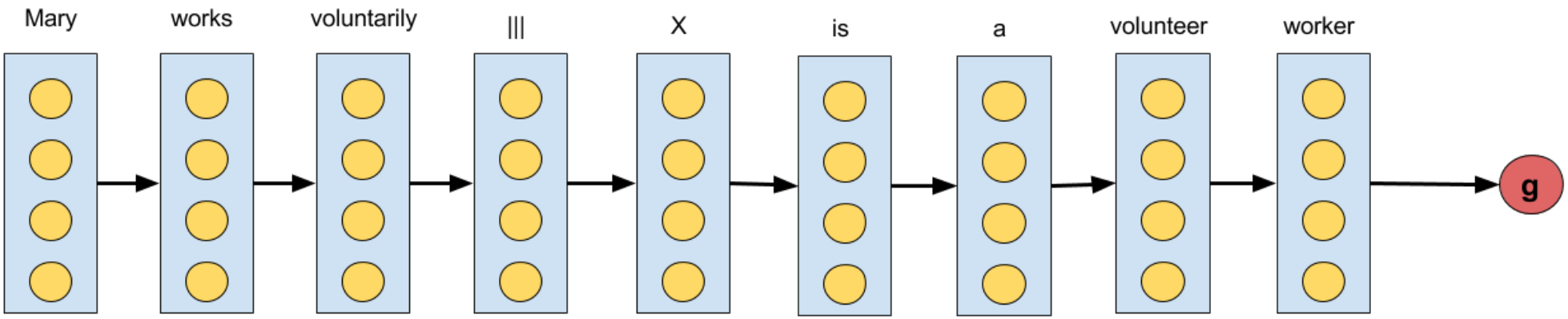
Vanishing Gradient Problem



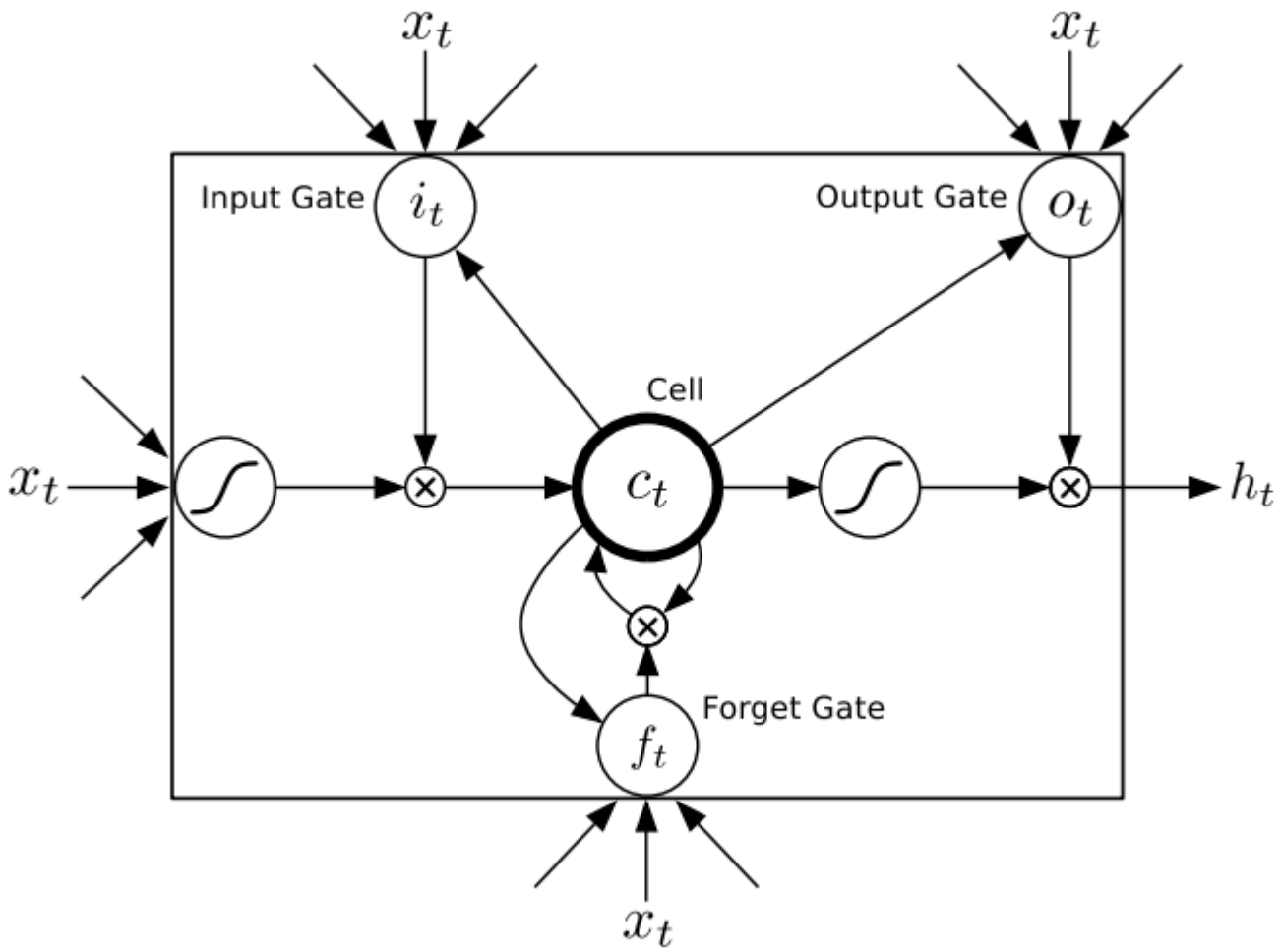
$$\nabla_{h^t} g = \frac{\partial_{h^{t+1}}}{\partial_{h^t}} \nabla_{h^{t+1}} g \ + \ \ldots$$

RNN

$$\vec{h}_{t+1} = tanh(W\vec{x}_{t+1} + U\vec{h}_t + b)$$



Long Short-Term Memory



[Hochreiter, Schmidhuber 1997; Graves 2013]

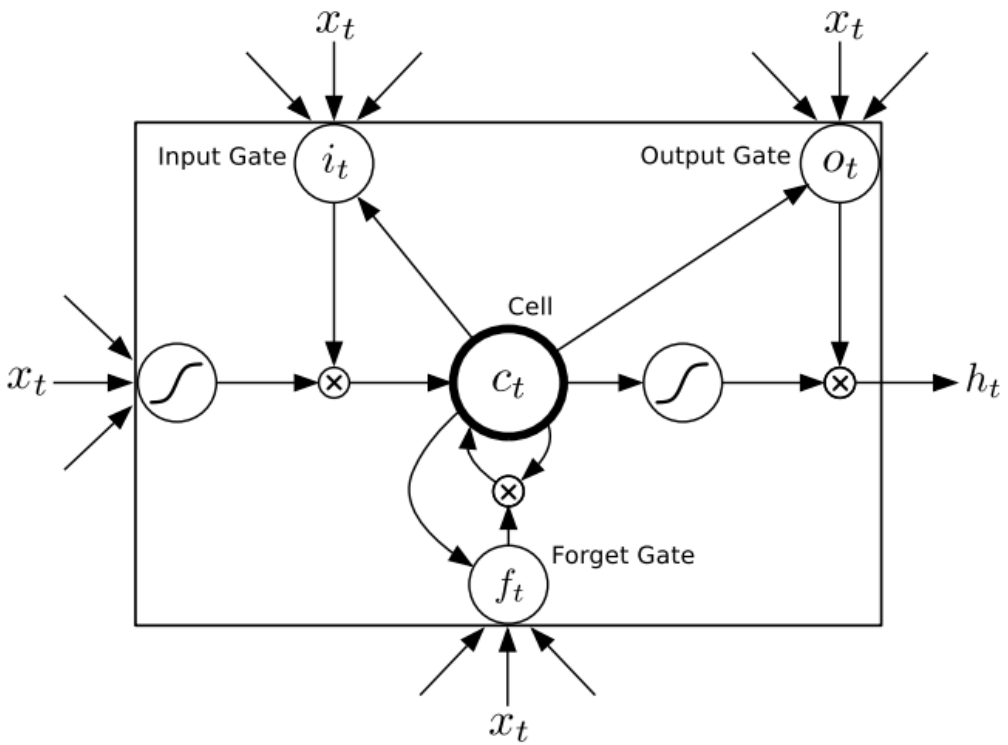
Long Short-Term Memory

$$i_t = \sigma\left(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i\right)$$ **Input gate**

$$f_t = \sigma\left(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f\right)$$ **Forget gate**

$$c_t = f_t c_{t-1} + i_t \tanh\left(W_{xc}x_t + W_{hc}h_{t-1} + b_c\right)$$

$$o_t = \sigma\left(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o\right)$$



$$h_t = o_t \tanh(c_t)$$

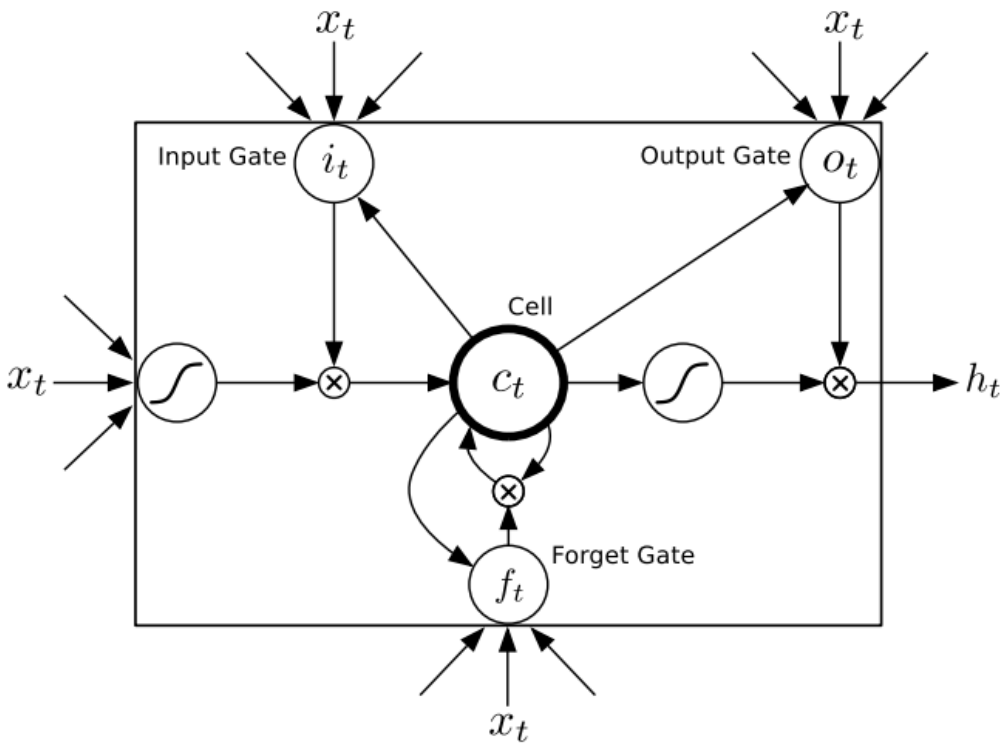
Long Short-Term Memory

$$i_t = \sigma\left(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i\right)$$ **Input gate**

$$f_t = \sigma\left(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f\right)$$ **Forget gate**

$$c_t = f_t c_{t-1} + i_t \tanh\left(W_{xc}x_t + W_{hc}h_{t-1} + b_c\right)$$ **Memory cell**

$$o_t = \sigma\left(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o\right)$$



$$h_t = o_t \tanh(c_t)$$

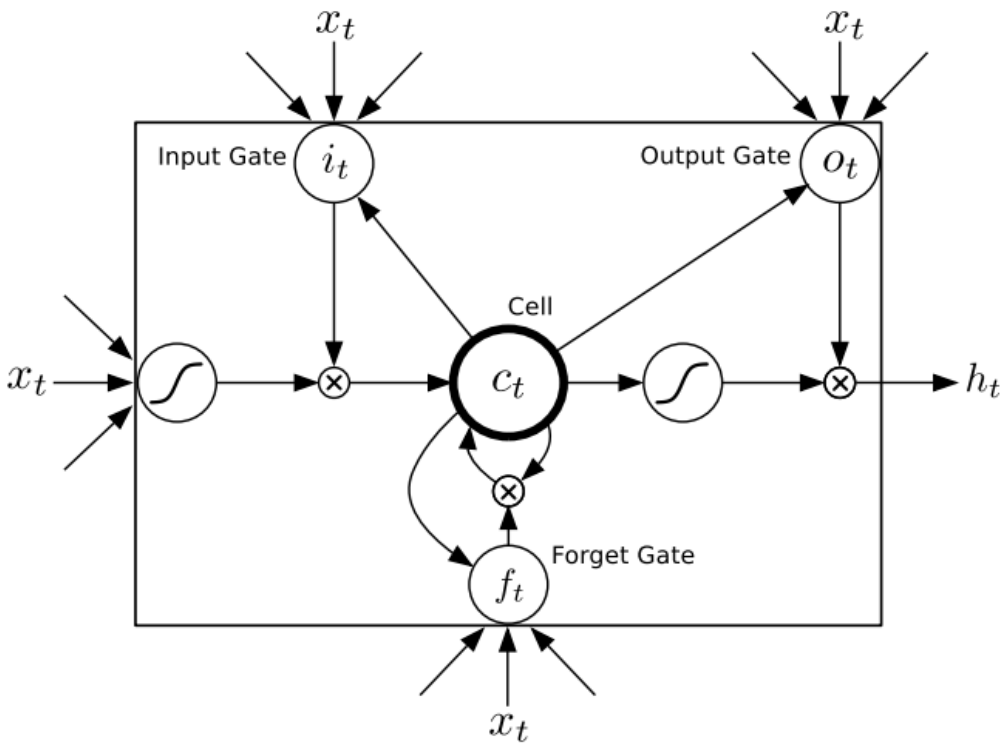
Long Short-Term Memory

$$i_t = \sigma\left(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i\right)$$ **Input gate**

$$f_t = \sigma\left(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f\right)$$ **Forget gate**

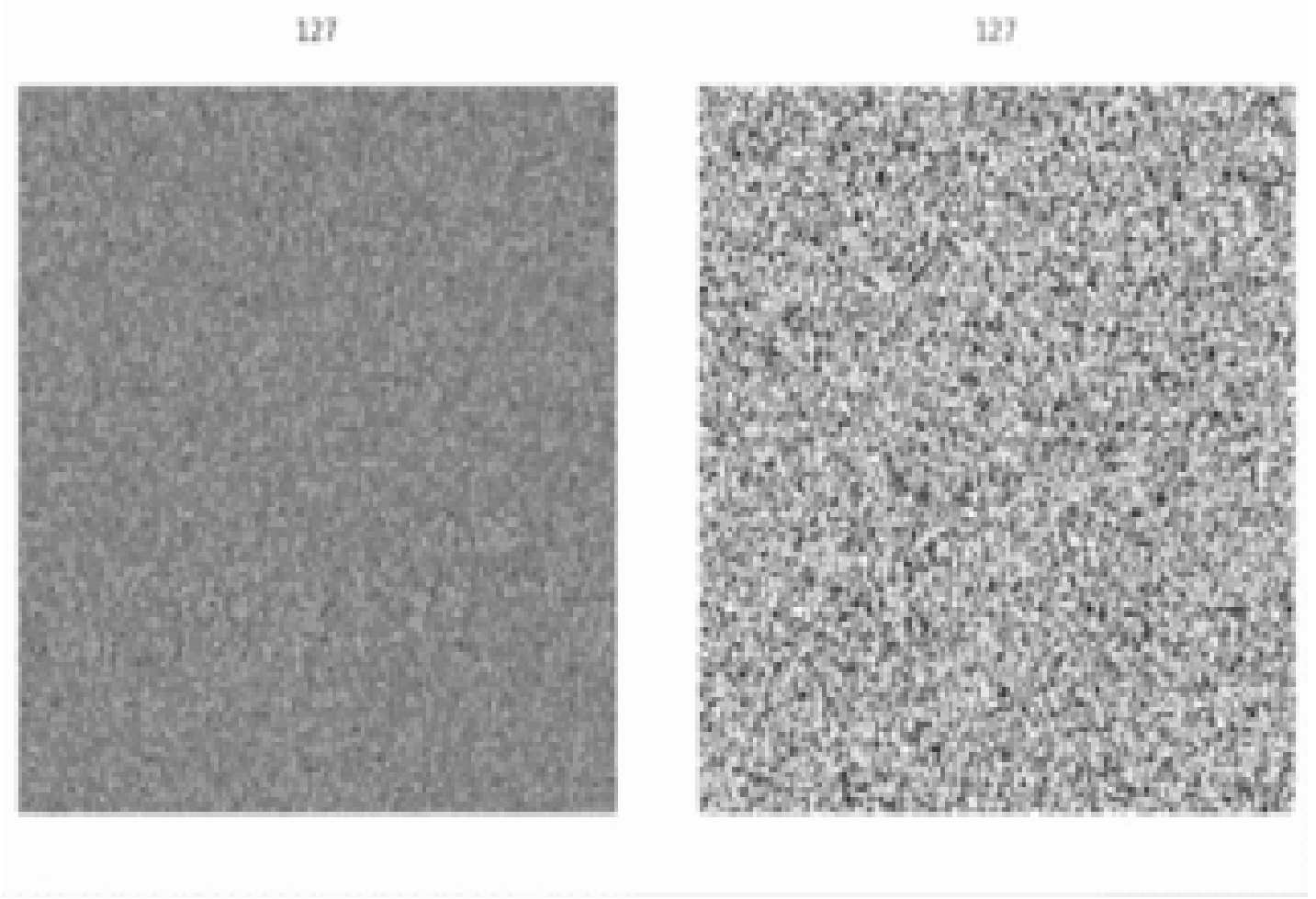
$$c_t = f_t c_{t-1} + i_t \tanh\left(W_{xc}x_t + W_{hc}h_{t-1} + b_c\right)$$ **Memory cell**

$$o_t = \sigma\left(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o\right)$$ **Output gate**

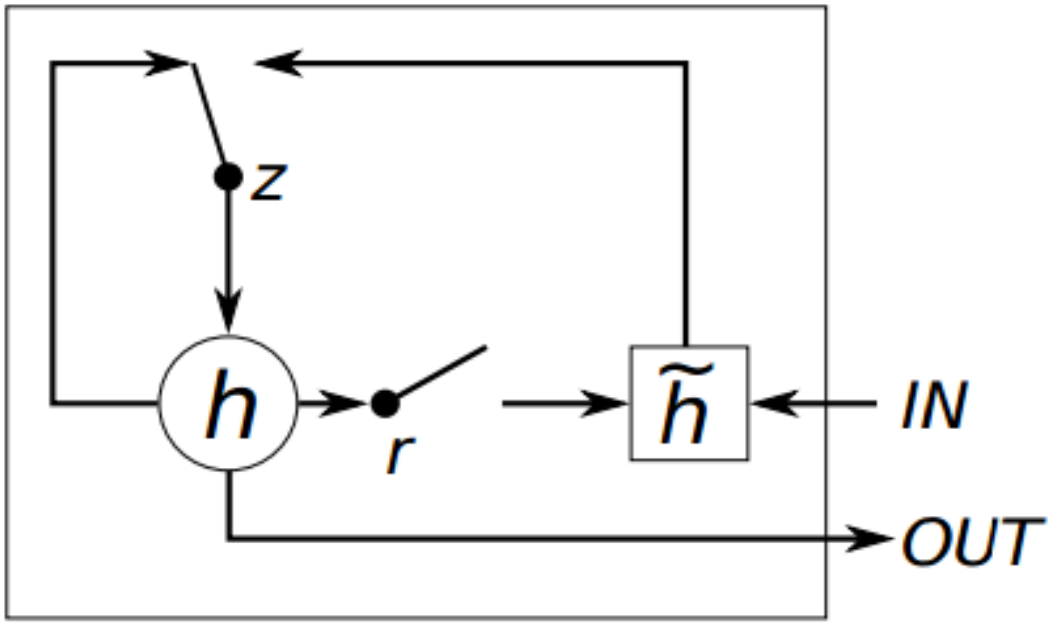


$$h_t = o_t \tanh(c_t)$$

Vanishing Gradients
RNN vs LSTM



Gated Recurrent Unit



[Chung 2014; Cho 2014]

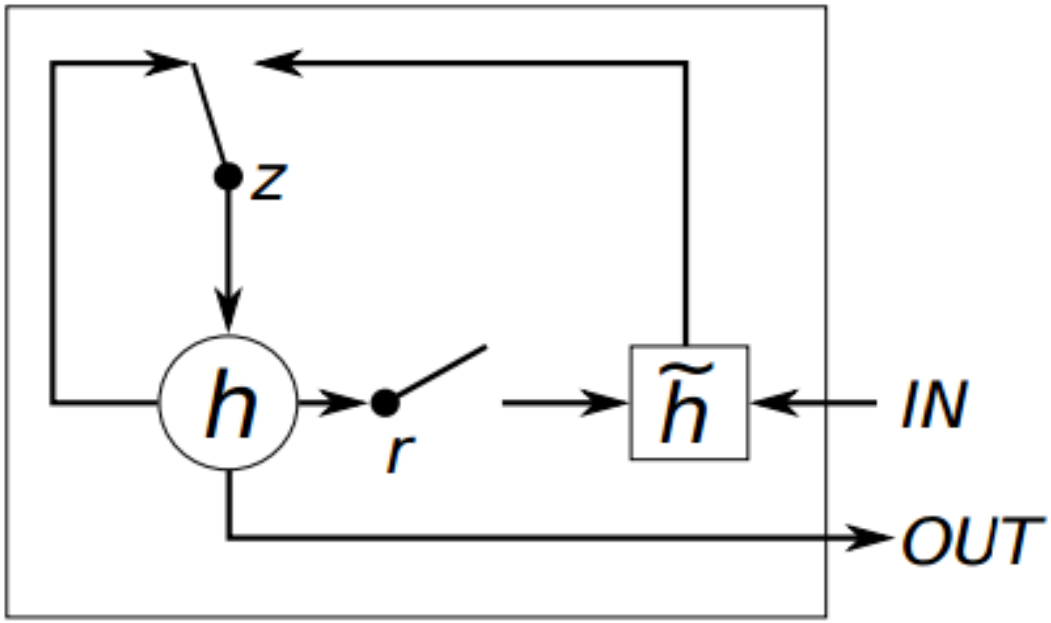
Gated Recurrent Unit

Update gate

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z)$$

Reset gate

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r)$$



$$\tilde{h}_t = \tanh(W x_t + U(r_t \odot h_{t-1}))$$

$$h_t = (1 - z_t) h_{t-1} + z_t \tilde{h}_t$$