

Deep Learning (CS324)

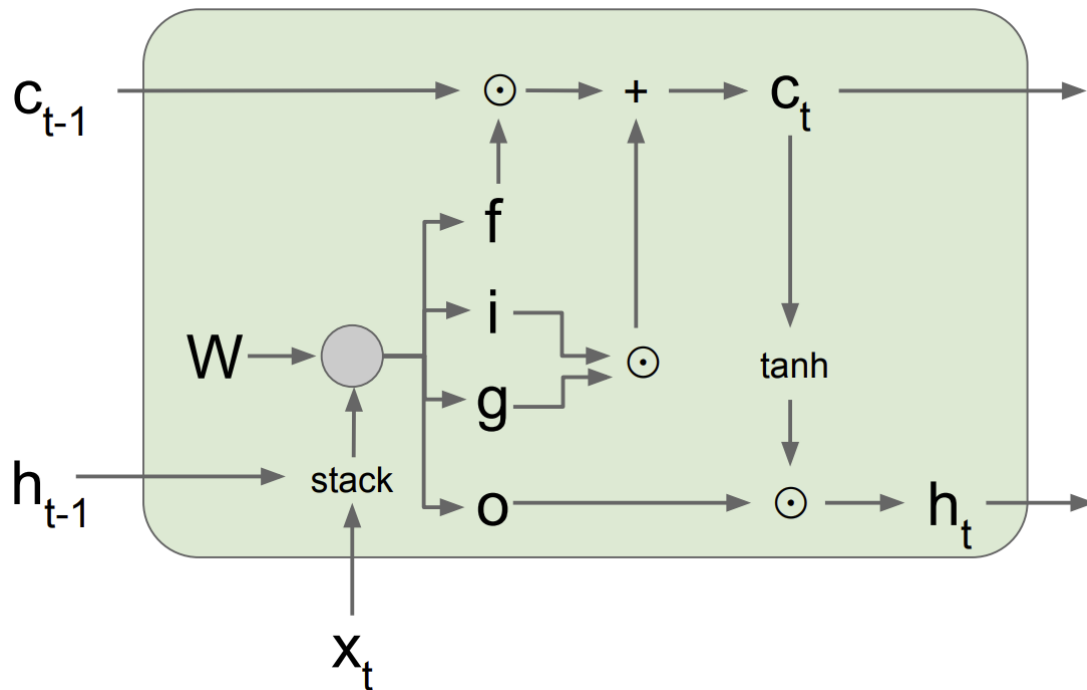
6. Recurrent Neural Networks* (Continued)

Jianguo Zhang
SUSTech

RNN

- Gated Recurrent Neural Network
- Different architectures of RNN.

LSTM Forward Pass Summary

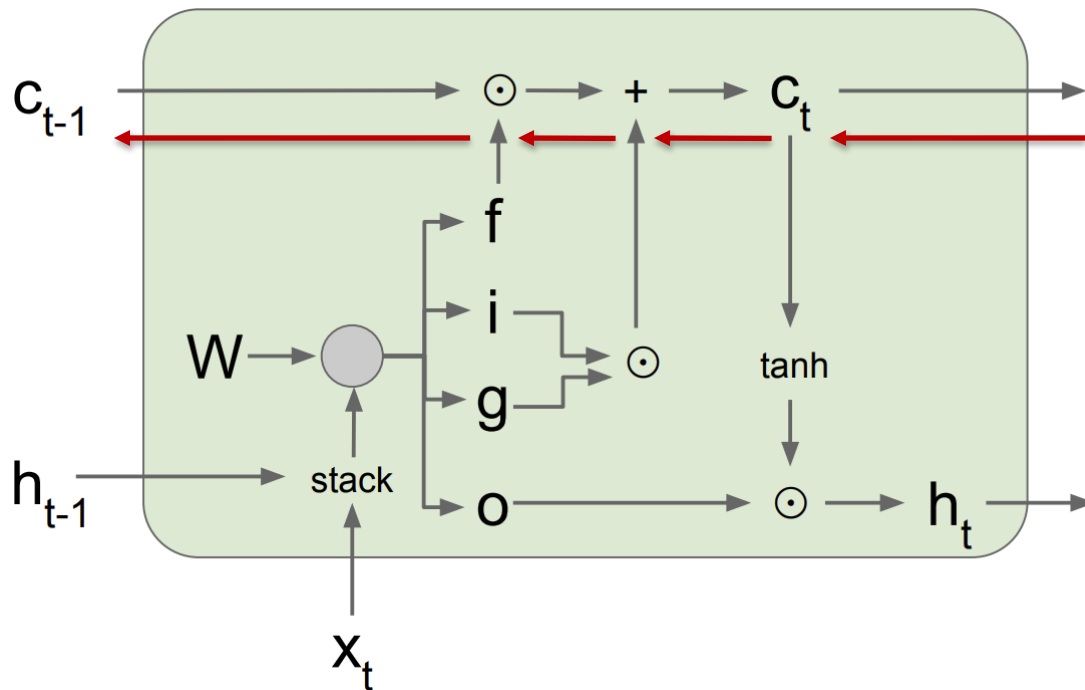


$$\begin{pmatrix} g_t \\ i_t \\ f_t \\ o_t \end{pmatrix} = \begin{pmatrix} \tanh \\ \sigma \\ \sigma \\ \sigma \end{pmatrix} \begin{pmatrix} W_g \\ W_i \\ W_f \\ W_o \end{pmatrix} \begin{pmatrix} x_t \\ h_{t-1} \end{pmatrix}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t$$

$$h_t = o_t \odot \tanh c_t$$

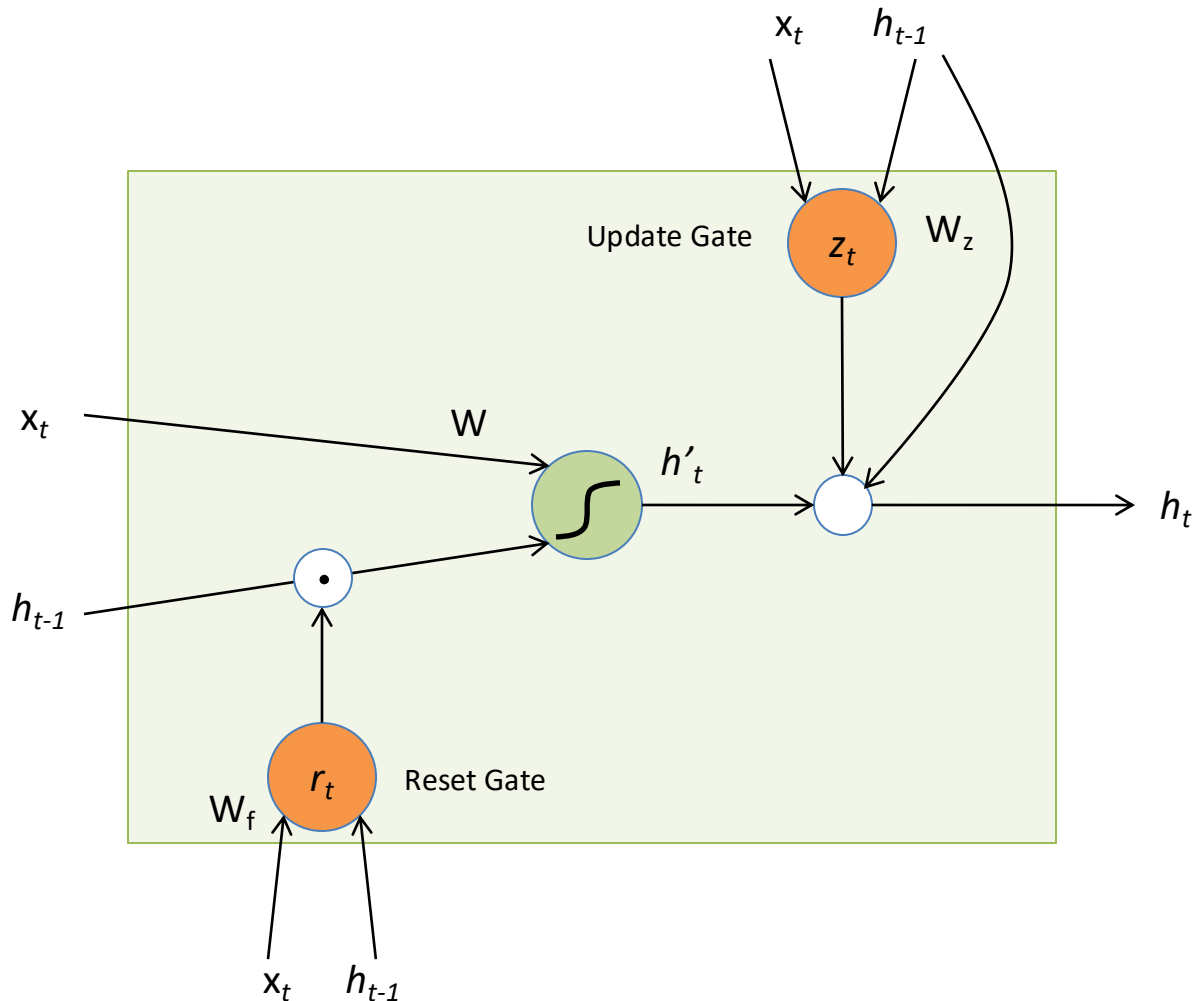
LSTM Backward Pass



Gradient flow from c_t to c_{t-1} only involves back-propagating through addition and elementwise multiplication, not matrix multiplication or tanh

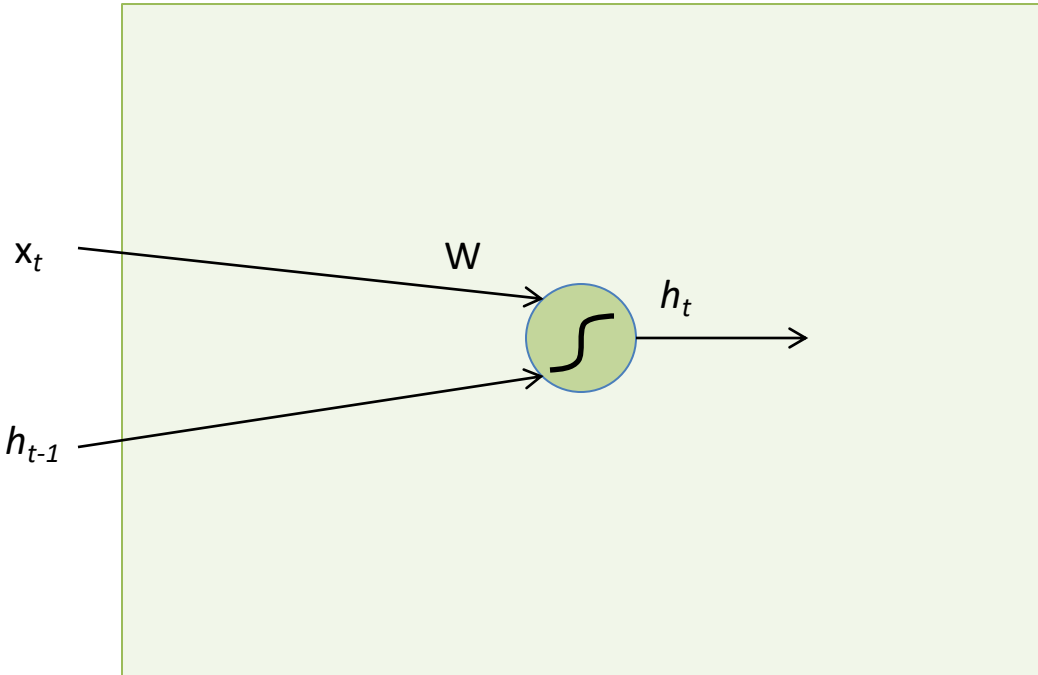
For complete details: [Illustrated LSTM Forward and Backward Pass](#)

Gated Recurrent Unit (GRU)



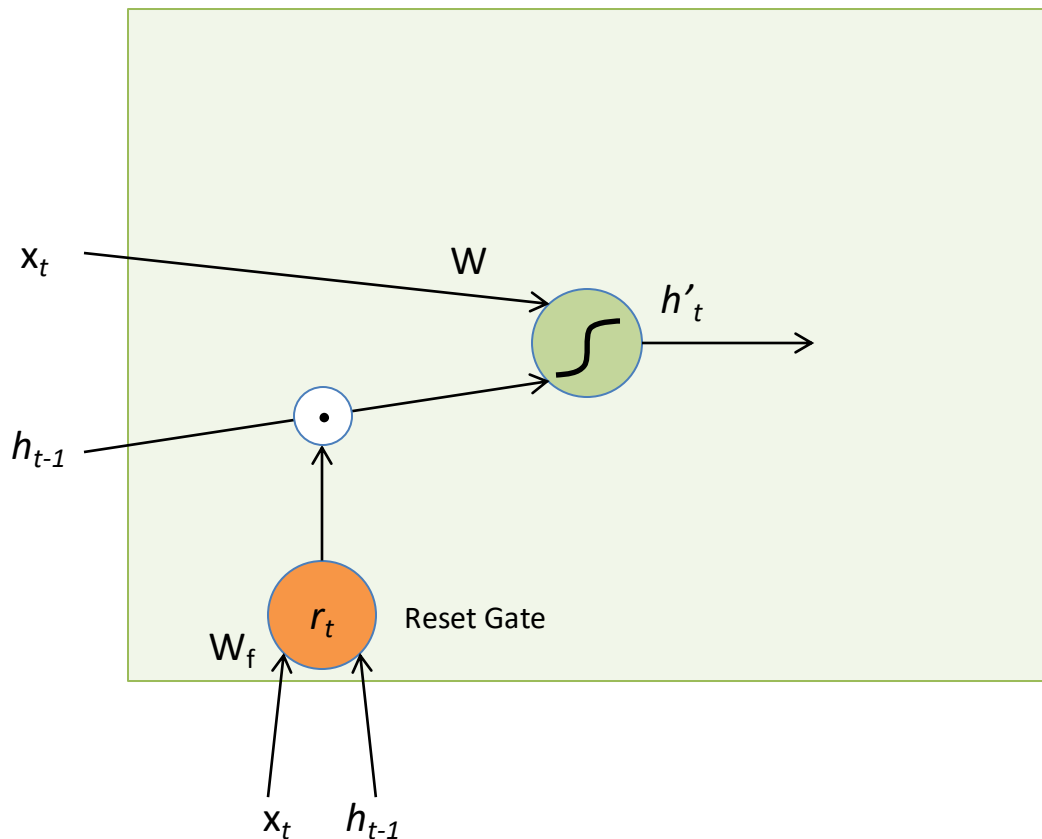
- Get rid of separate cell state
- Merge “forget” and “output” gates into “update” gate

Gated Recurrent Unit (GRU)



$$h_t = \tanh W \begin{pmatrix} x_t \\ h_{t-1} \end{pmatrix}$$

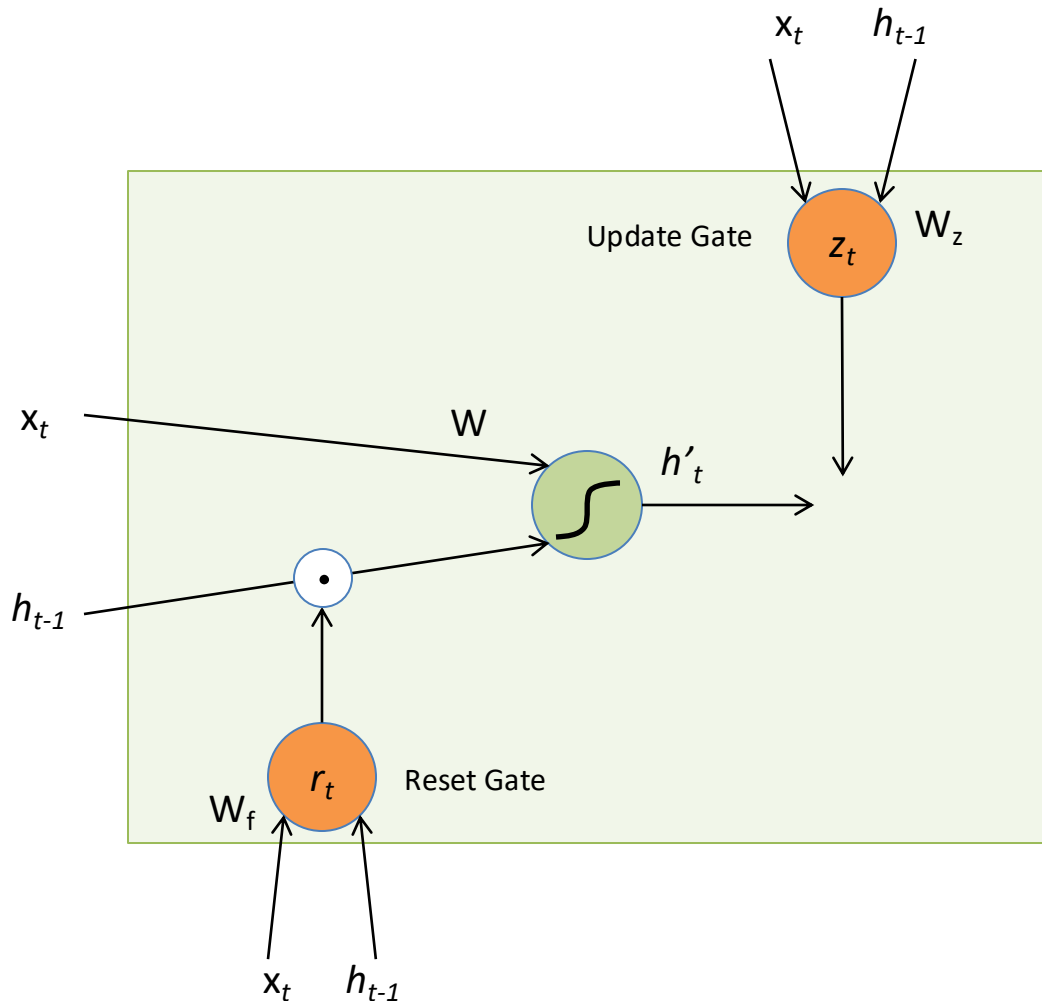
Gated Recurrent Unit (GRU)



$$r_t = \sigma \left(W_r \begin{pmatrix} x_t \\ h_{t-1} \end{pmatrix} + b_t \right)$$

$$h'_t = \tanh W \left(r_t \odot h_{t-1} \right)$$

Gated Recurrent Unit (GRU)

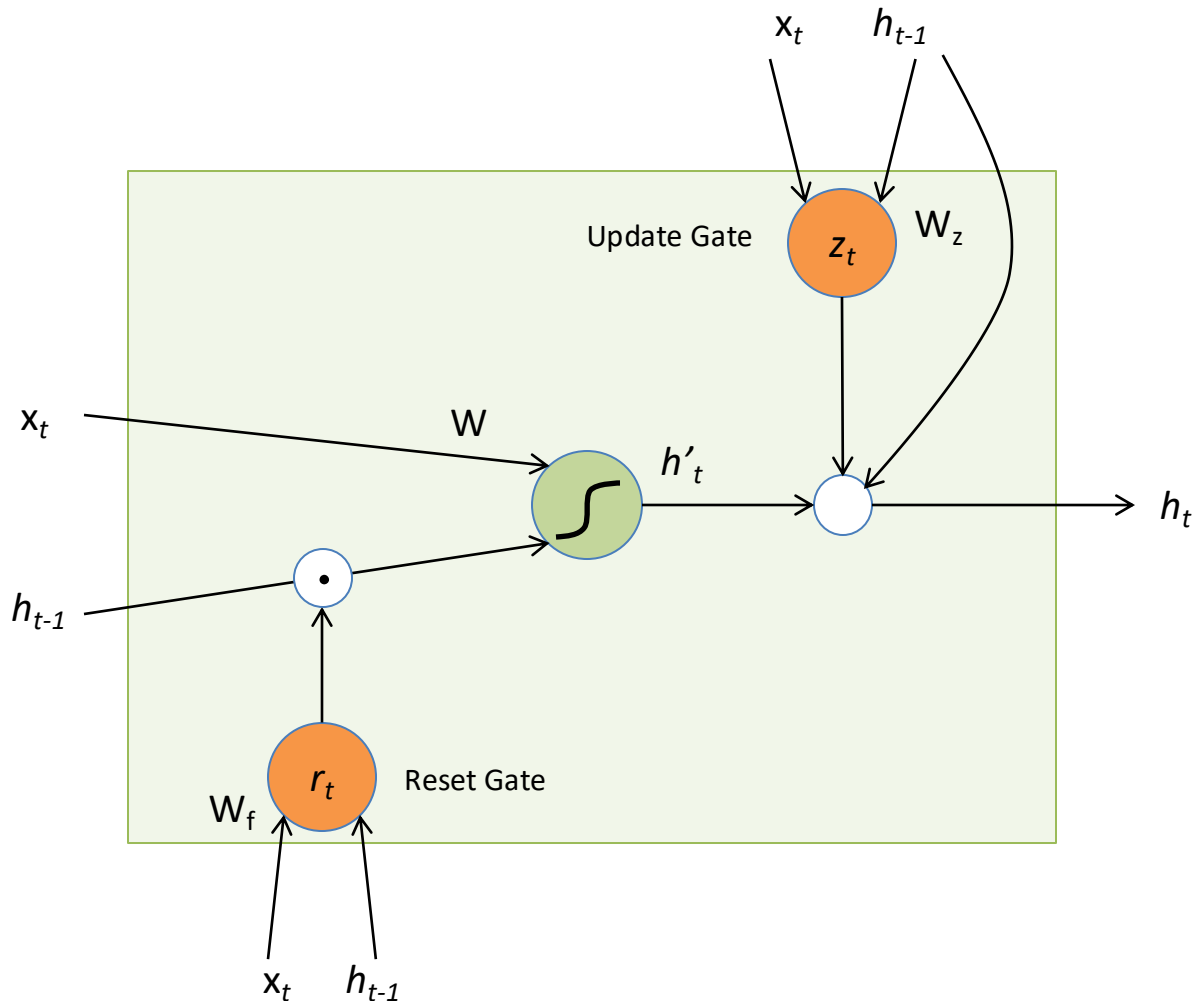


$$r_t = \sigma \left(W_r \begin{pmatrix} x_t \\ h_{t-1} \end{pmatrix} + b_r \right)$$

$$h'_t = \tanh W \begin{pmatrix} x_t \\ r_t \odot h_{t-1} \end{pmatrix}$$

$$z_t = \sigma \left(W_z \begin{pmatrix} x_t \\ h_{t-1} \end{pmatrix} + b_z \right)$$

Gated Recurrent Unit (GRU)



$$r_t = \sigma \left(W_r \begin{pmatrix} x_t \\ h_{t-1} \end{pmatrix} + b_r \right)$$

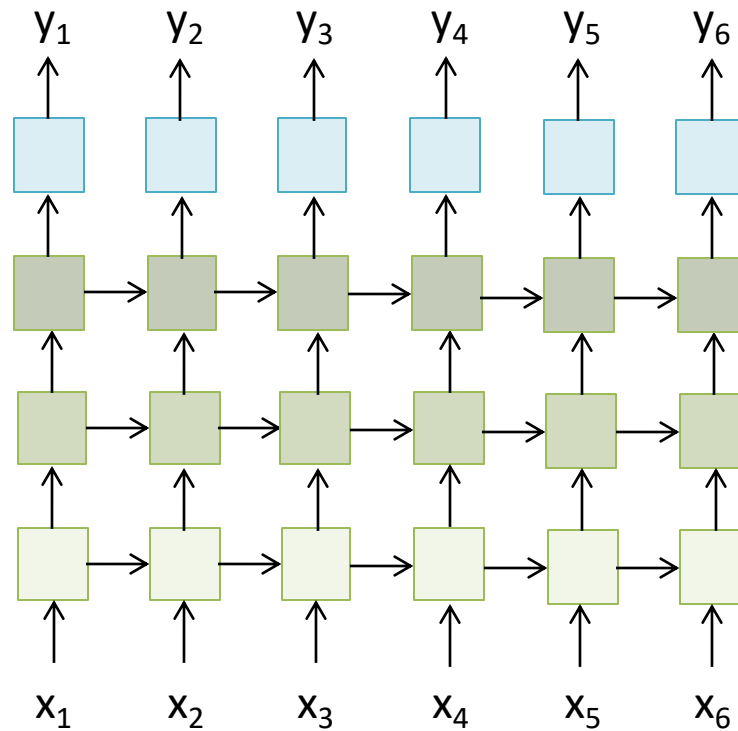
$$h'_t = \tanh W \begin{pmatrix} x_t \\ r_t \odot h_{t-1} \end{pmatrix}$$

$$z_t = \sigma \left(W_z \begin{pmatrix} x_t \\ h_{t-1} \end{pmatrix} + b_z \right)$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot h'_t$$

Multi-layer RNNs

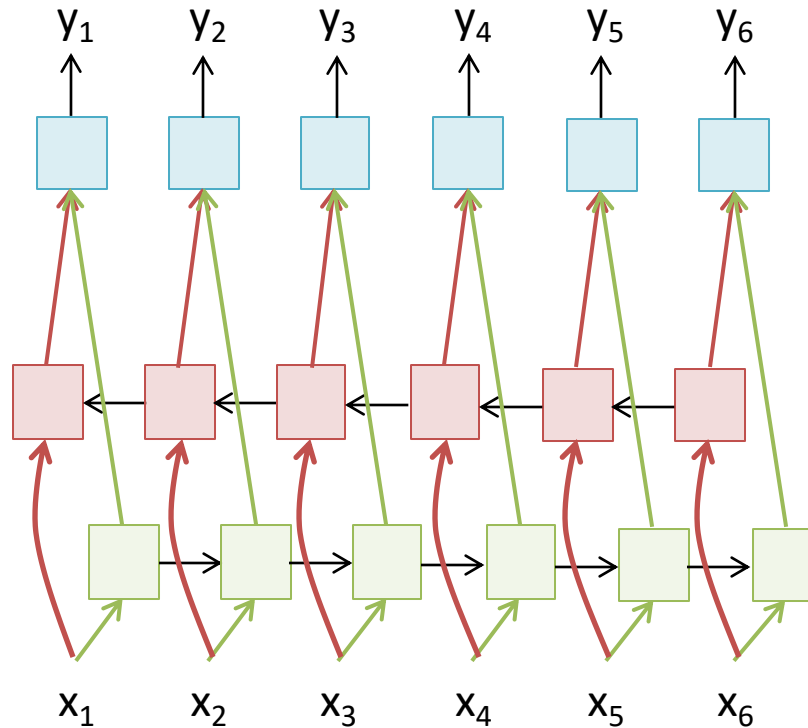
- We can of course design RNNs with multiple hidden layers



- Anything goes: skip connections across layers, across time, ...

Bi-directional RNNs

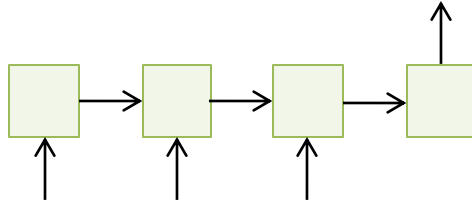
- RNNs can process the input sequence in forward and in the reverse direction



- Popular in speech recognition

Use Cases

Multiple input –
Single output



Sequence Classification

Single - Multiple

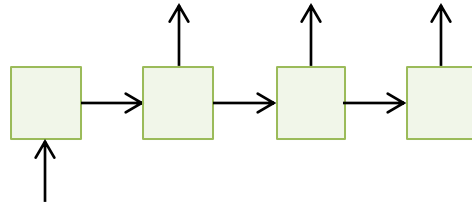


Image Captioning

Multiple - Multiple

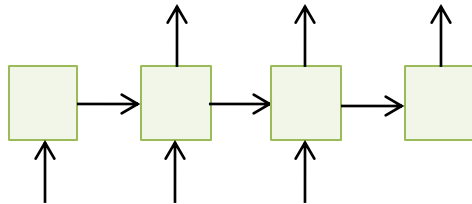
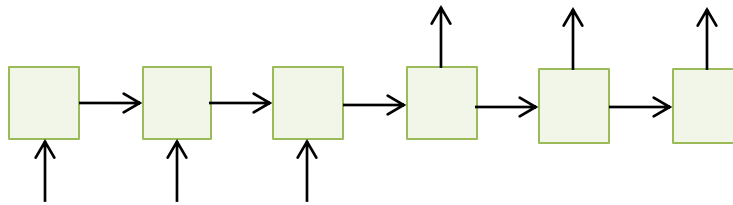


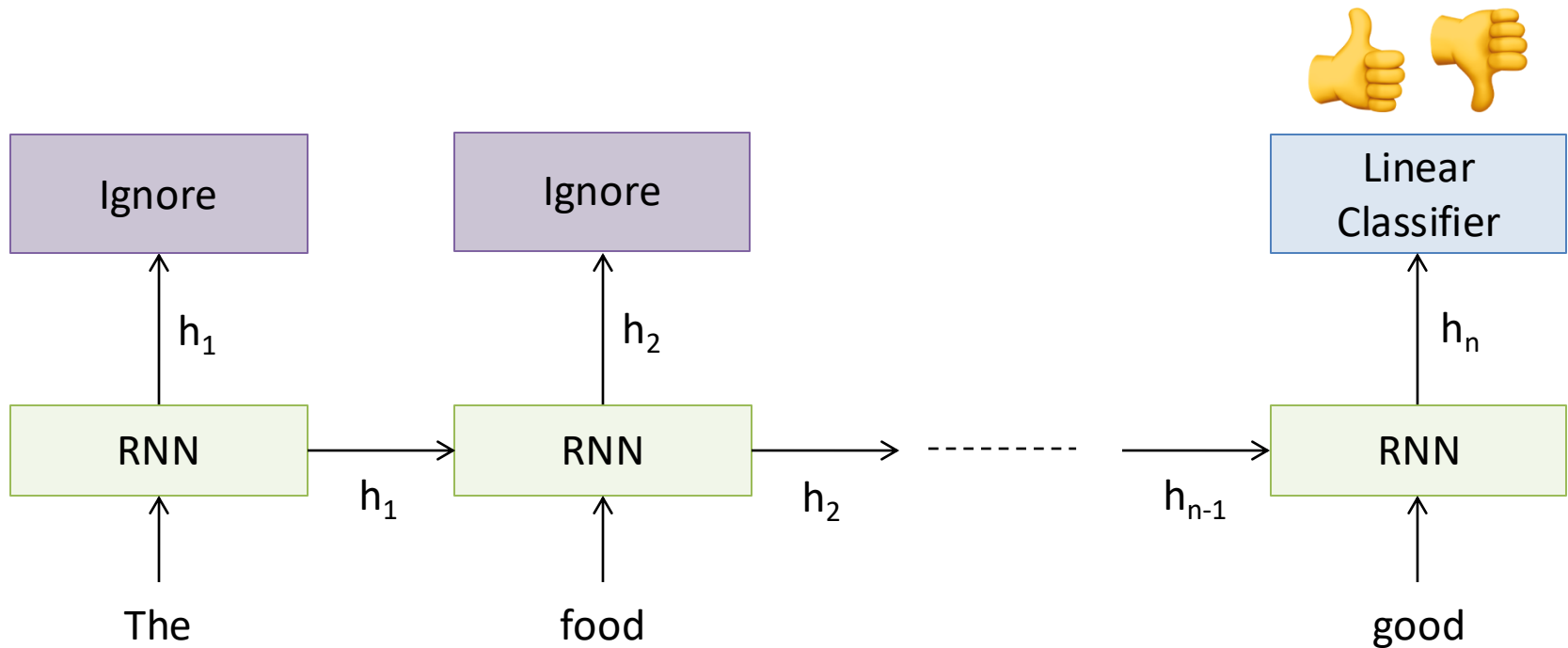
Image Captioning

Multiple - Multiple

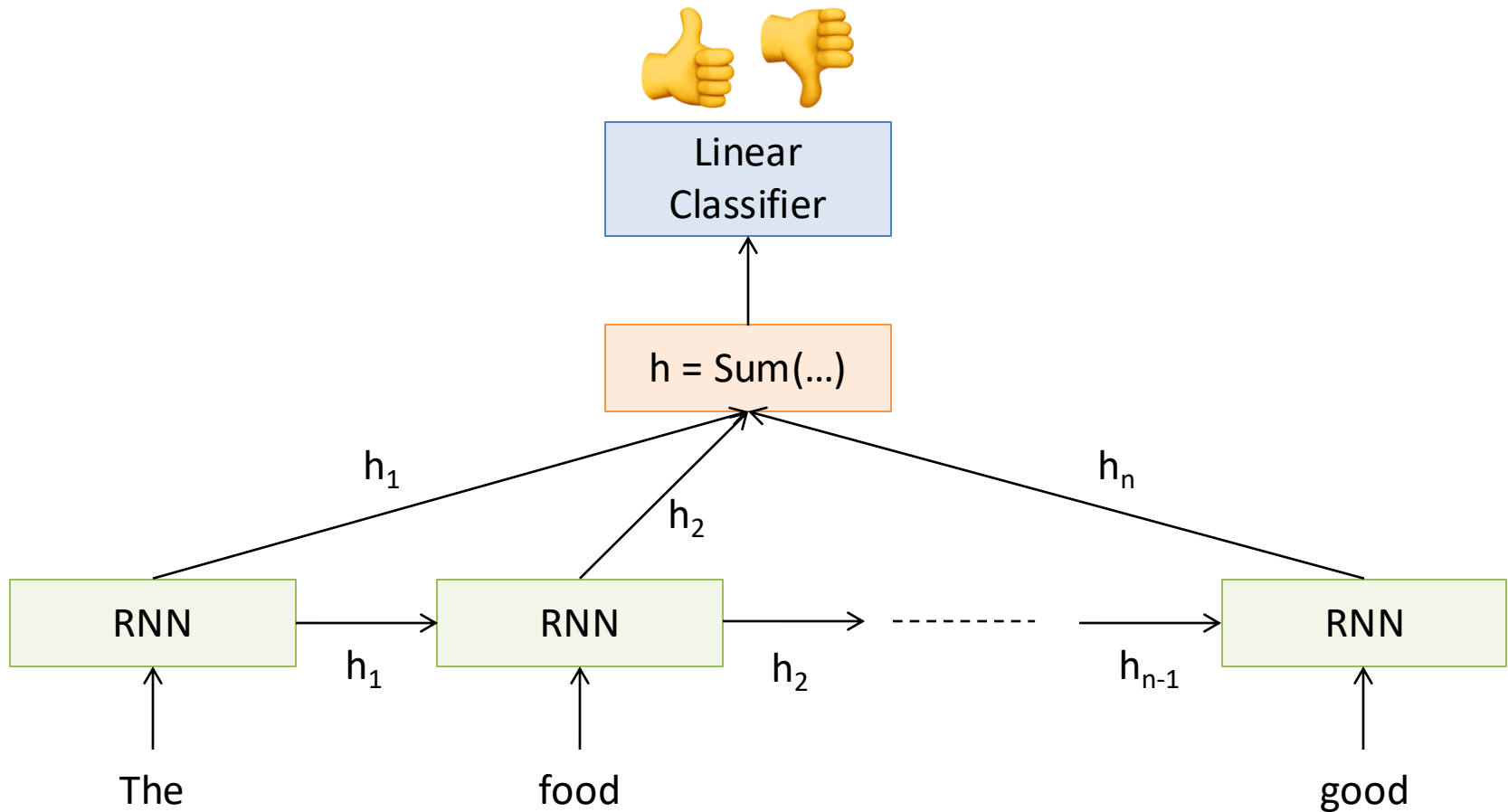


Translation

Sequence Classification



Sequence Classification



Sequence Classification

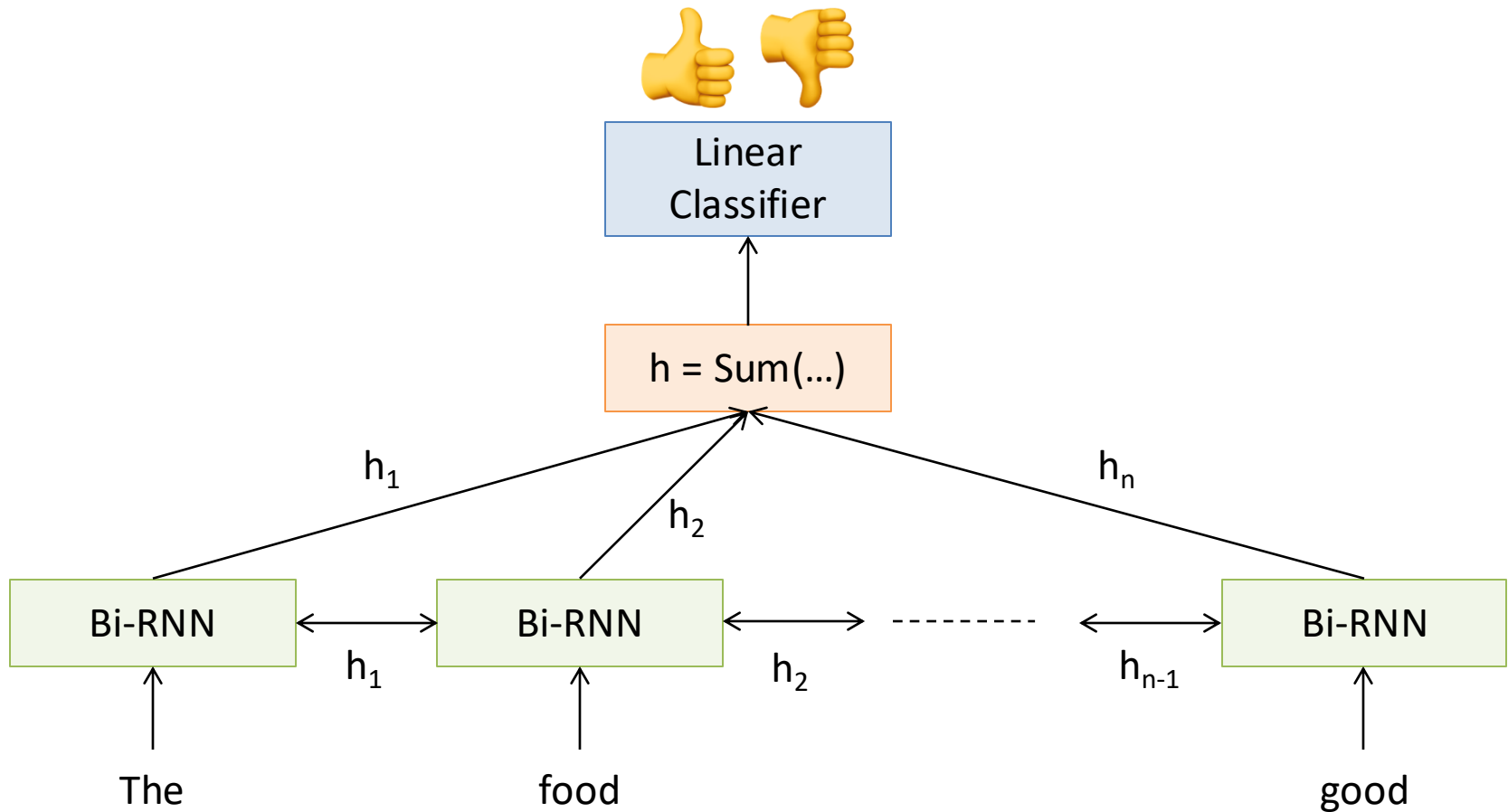
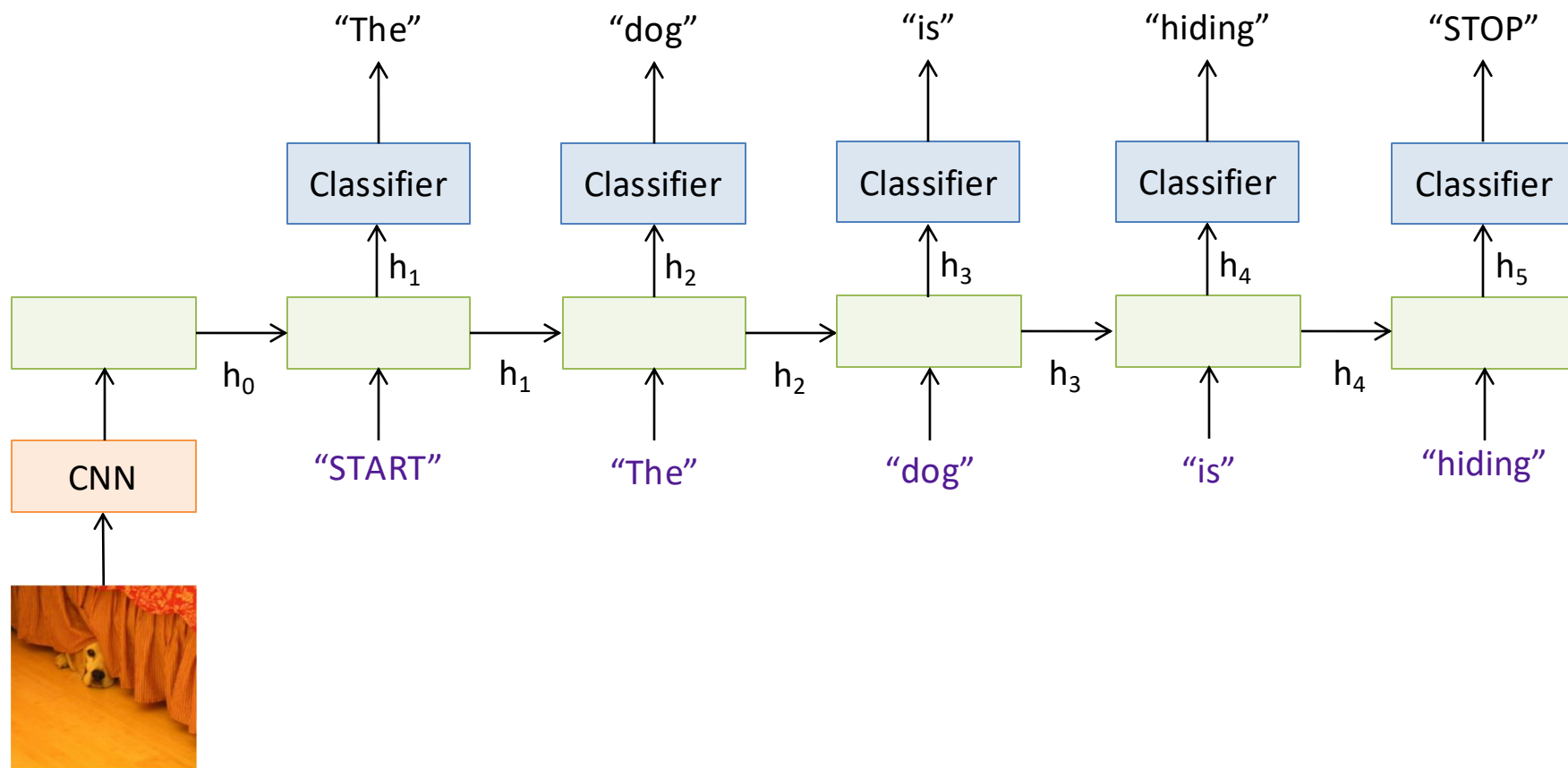


Image Caption Generation



It's raining LSTMs

- There exist countless variations of LSTMs, with different researchers proposing different arrangements of the LSTM units
- So, which one is better?
- **None:** <https://arxiv.org/pdf/1503.04069.pdf>
- Also, **RNNs can outperform both LSTMs and GRUs:** <http://proceedings.mlr.press/v37/jozefowicz15.pdf>

A Zoo of RNNs

- If you are interested in the details for the following topics:
 - Bidirectional RNNs (Sec. 10.3 book)
 - Teacher forcing (Fig. 10.6 book)
 - Image captioning RNNs (Fig. 10.9 book)
 - Encoder-decoder architectures (Sec. 10.4 book)
 - ...

Online resources

- Music composition:
<http://www.hexahedria.com/2015/08/03/composing-music-with-recurrent-neural-networks/>
- Characters prediction:
<https://cs.stanford.edu/people/karpathy/recurrentjs/>
- Transformer networks:
<https://ai.googleblog.com/2017/08/transformer-novel-neural-network.html>
- RNN-Ts on your phone:
<https://ai.googleblog.com/2019/03/an-all-neural-on-device-speech.html>

Summary

- Sequential data and temporal dependances
- Recurrent Neural Network and BPTT
- Long Short-Term Memory
- Gated Recurrent Unit
- Different application cases of RNN