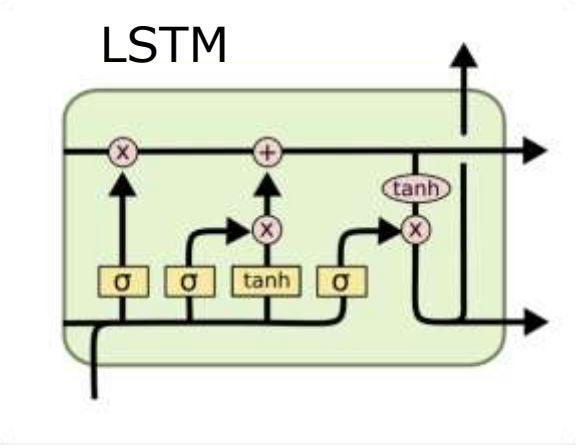
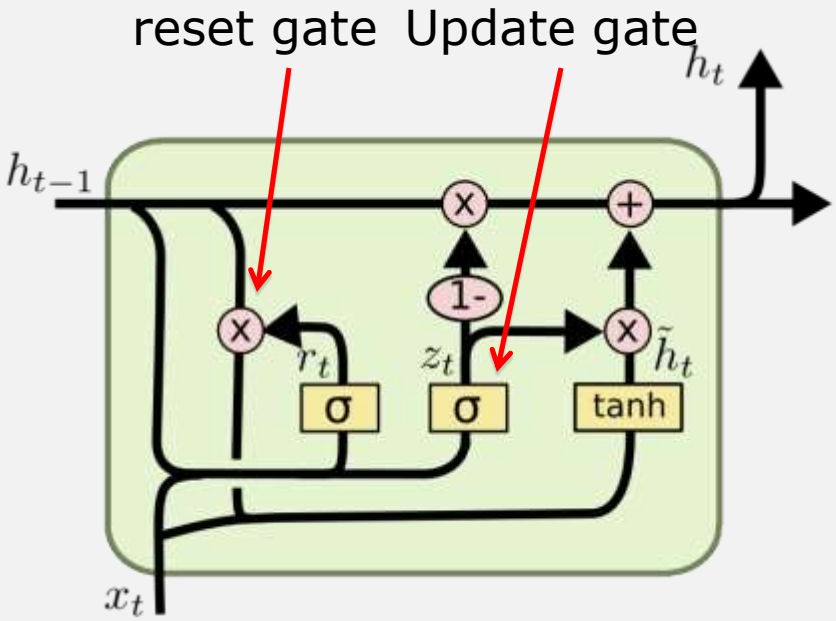
THE NORTHCAP UNIVERSITY
POWERED BY Arizona State University

Gated Recurrent Units(GRU)

GRU – Gated Recurrent Unit

(more compression)





$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right)$$

$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t]\right)$$

$$\tilde{h}_t = \tanh\left(W \cdot [r_t * h_{t-1}, x_t]\right)$$

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

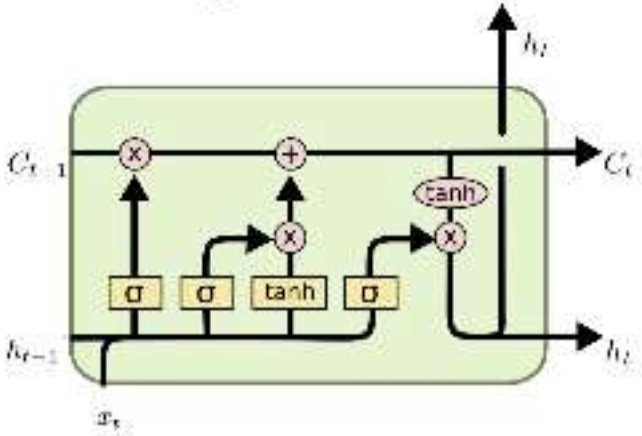
It combines the forget and input into a single update gate.

It also merges the cell state and hidden state. This is simpler

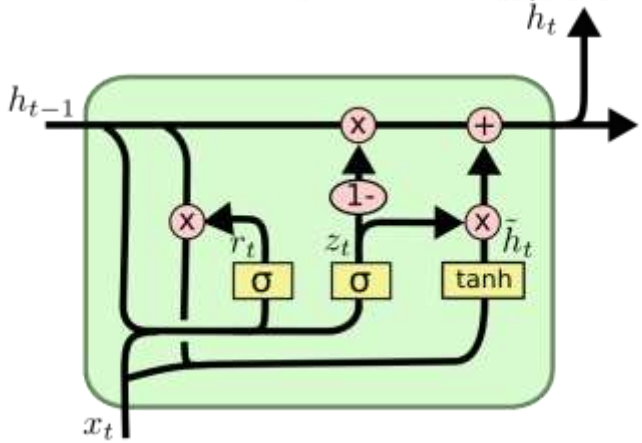
than LSTM. There are many other variants too. X,*: element-wise multi

LSTM and GRU

- LSTM [Hochreiter&Schmidhuber97]



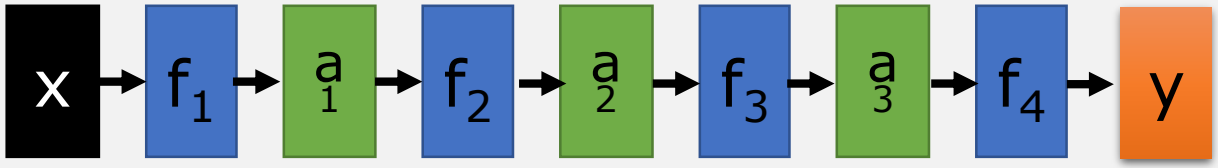
- GRU [Cho+14]



GRUs also takes x_t and h_{t-1} as inputs. They perform some calculations and then pass along h_t. What makes them different from LSTMs is that GRUs don't need the cell layer to pass values along. The calculations within each iteration insure that the h_t values being passed along either retain a high amount of old information or are jump-started with a high amount of new information.

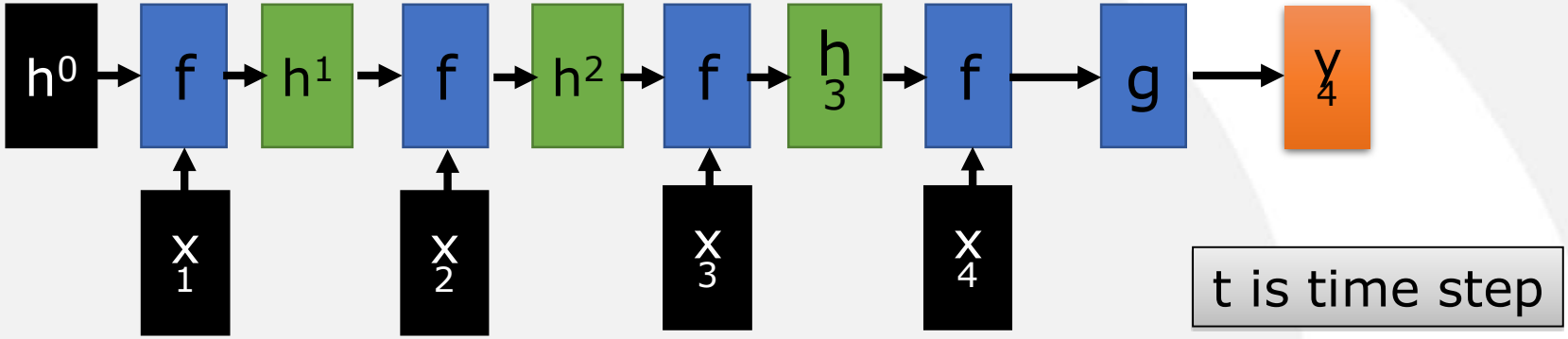
Feed-forward vs Recurrent Network

1. Feedforward network does not have input at each step
2. Feedforward network has different parameters for each layer



$a^t = f_t(a^{t-1}) = \sigma(W^t a^{t-1} + b^t)$

t is layer



t is time step

$a^t = f(a^{t-1}, x^t) = \sigma(W^h a^{t-1} + W^i x^t + b^i)$

We will turn the recurrent network 90 degrees.

GRU → Highway Network

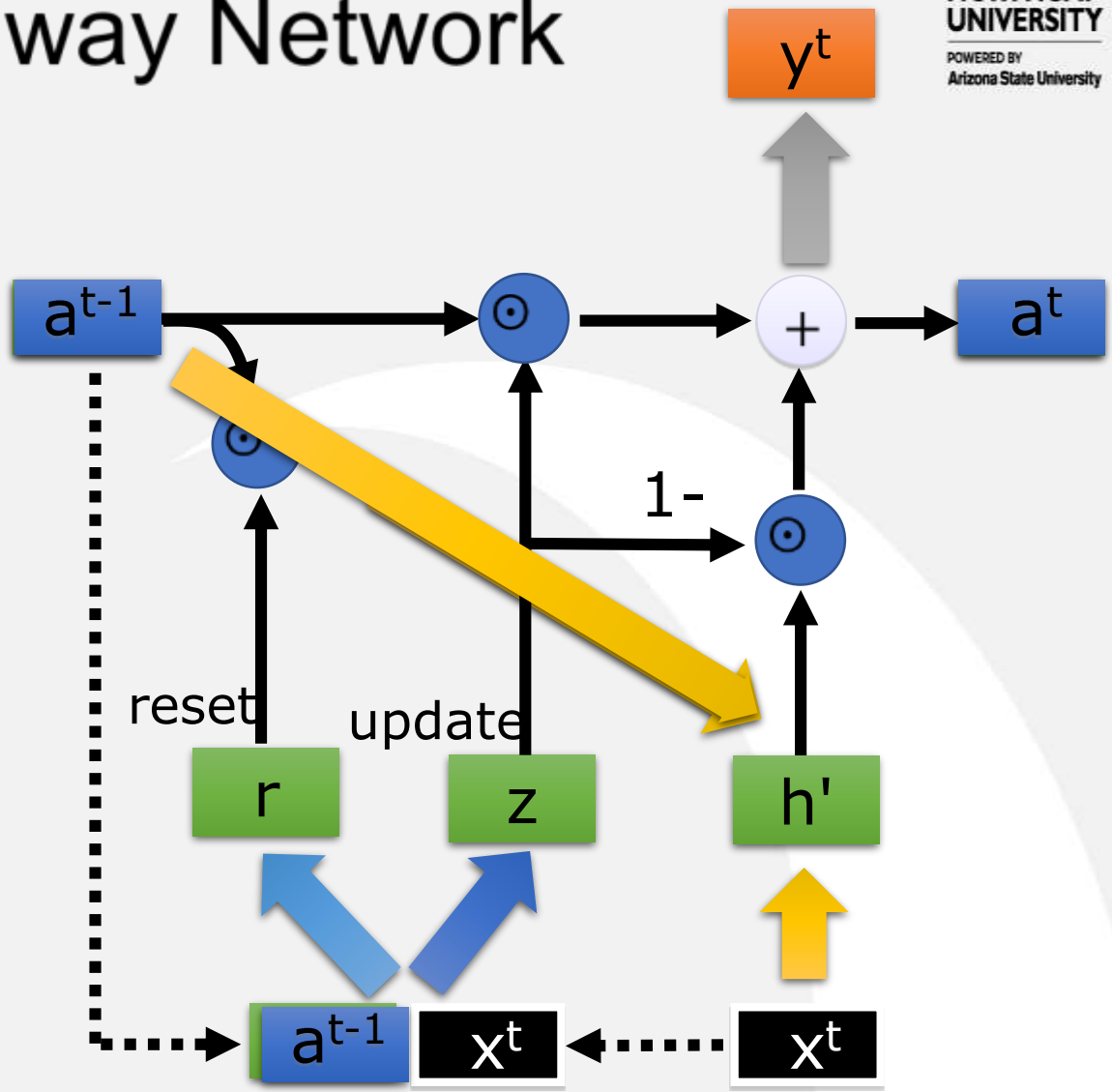
No input x^t at each step

No output y^t at each step

a^{t-1} is the output of the (t-1)-th layer

a^t is the output of the t-th layer

No reset gate



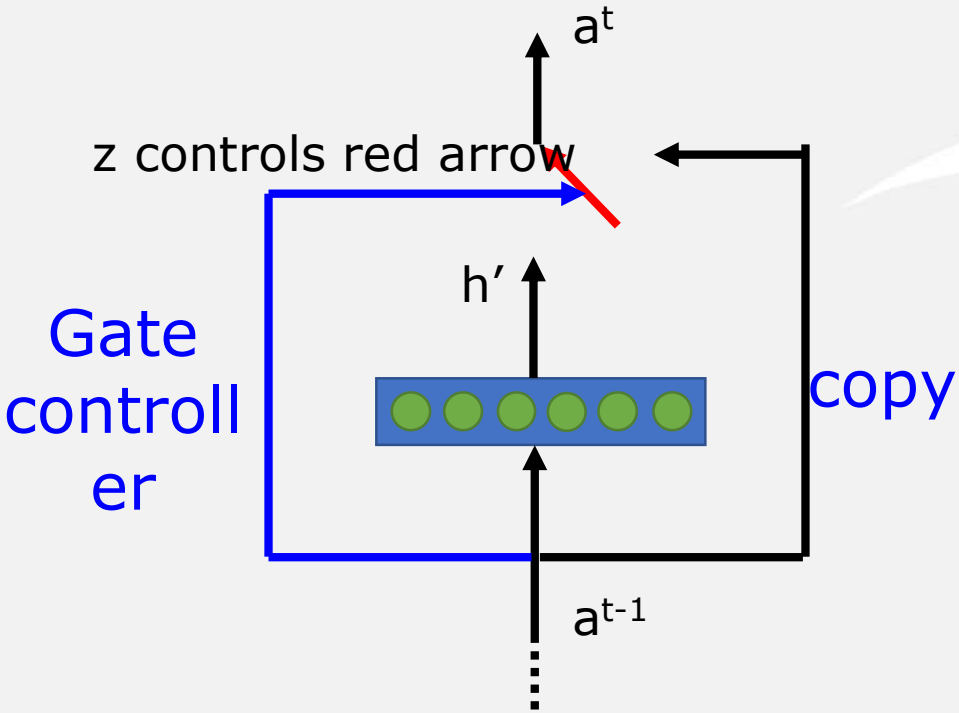
Highway Network

$$h' = \sigma(Wa^{t-1})$$

$$z = \sigma(W'a^{t-1})$$

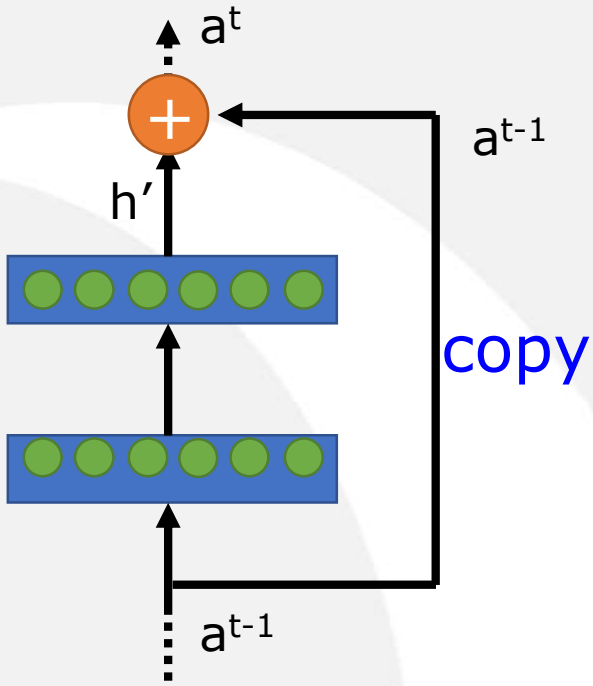
$$a^t = z \odot a^{t-1} + (1-z) \odot h'$$

- **Highway Network**

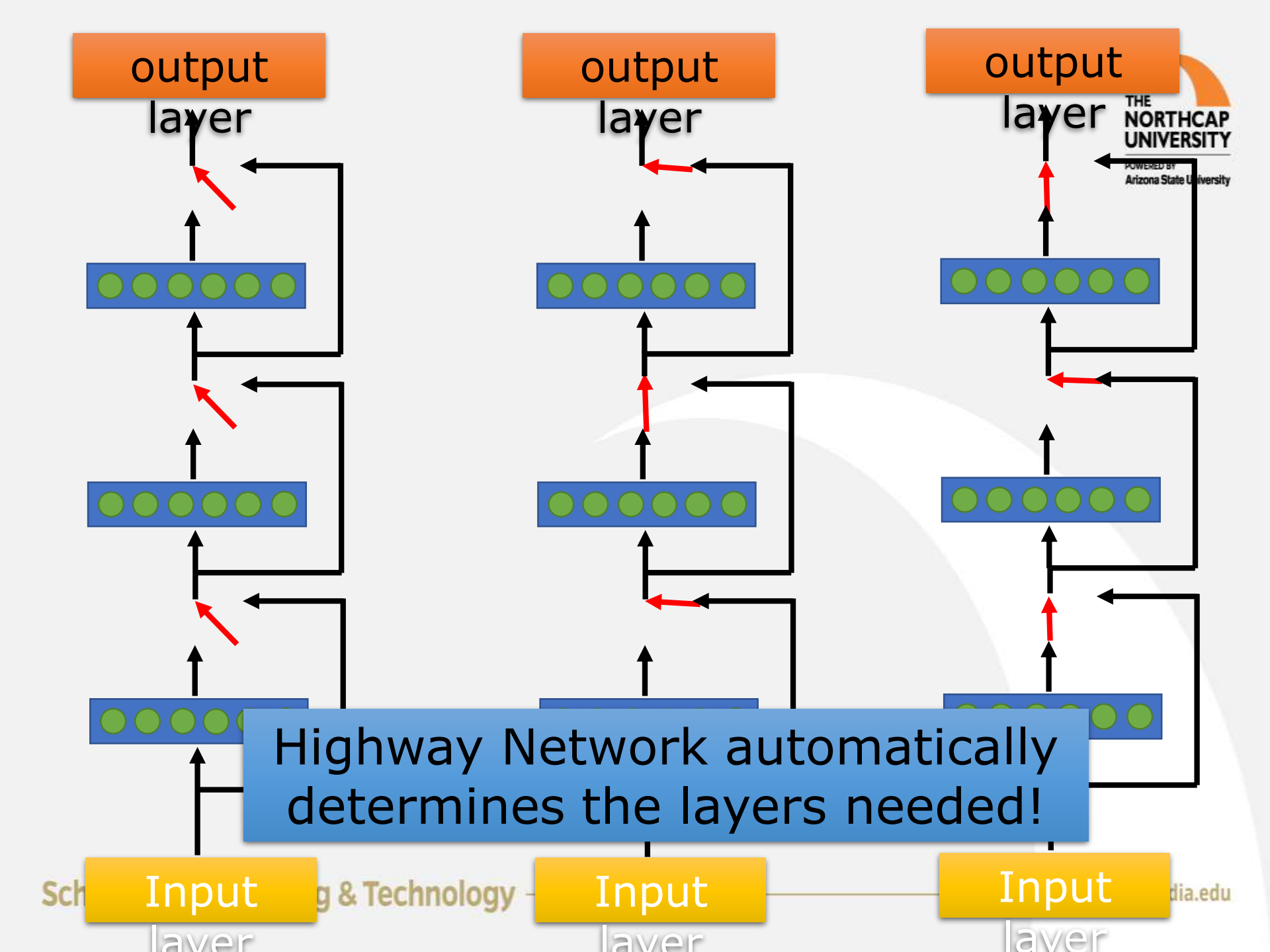


Training Very Deep Networks
https://arxiv.org/pdf/1507.06228v2.pdf

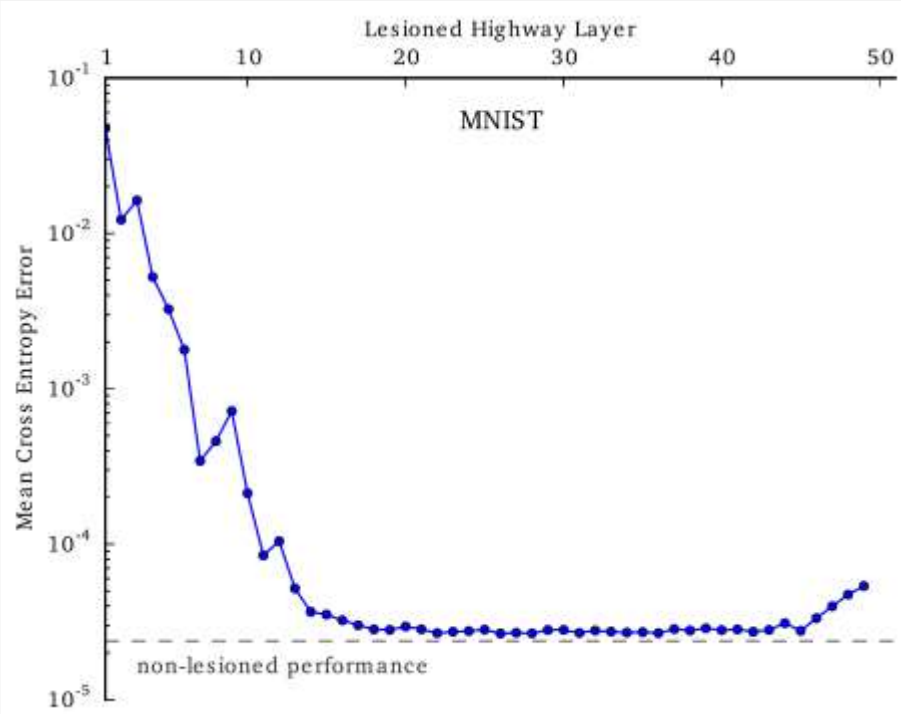
- **Residual Network**

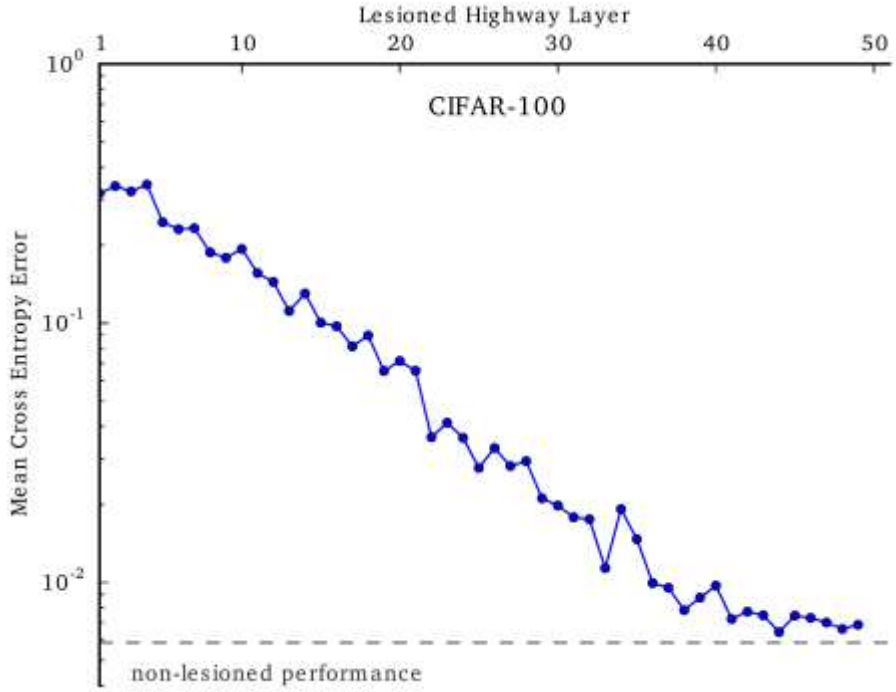


Deep Residual Learning for Image Recognition
http://arxiv.org/abs/1512.03385


output layer
output layer
output layer
THE NORTHCAP UNIVERSITY
POWERED BY Arizona State University
Highway Network automatically determines the layers needed!
Input layer
Input layer
Input layer

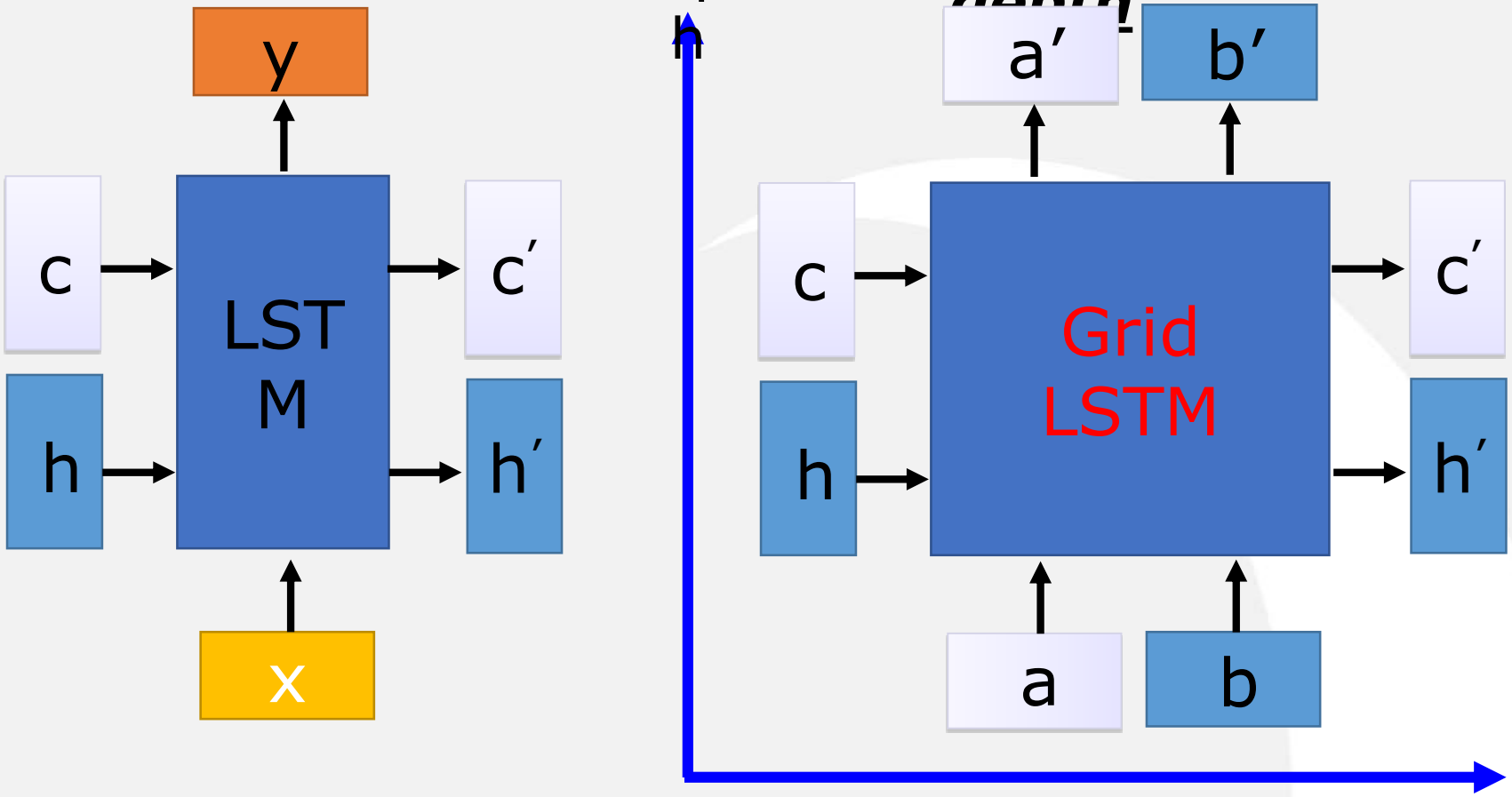
Highway Network Experiments



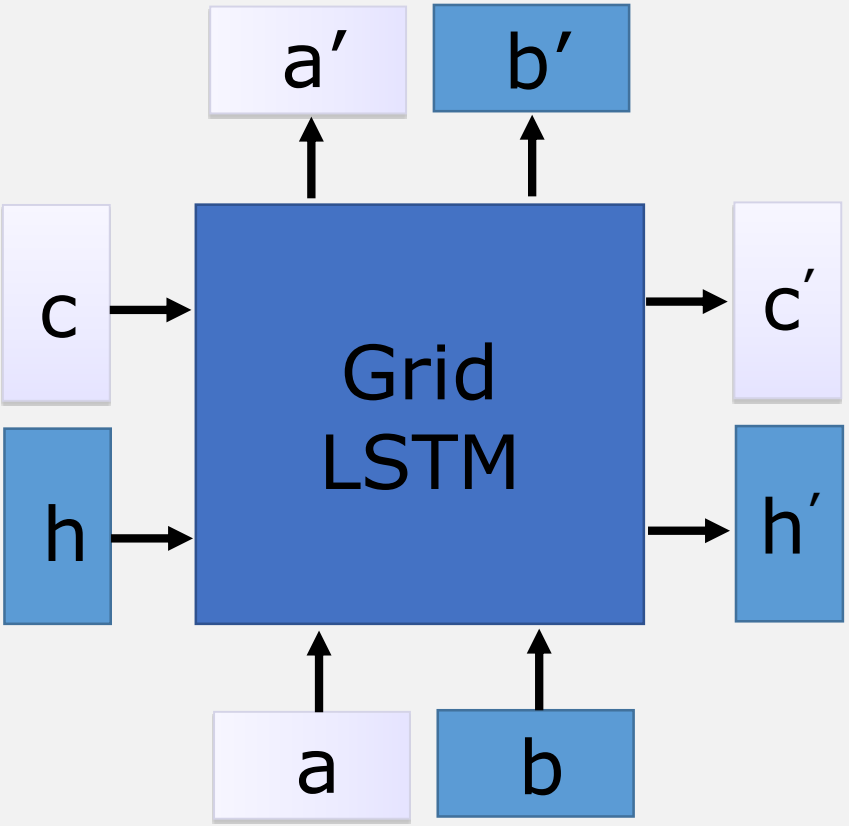


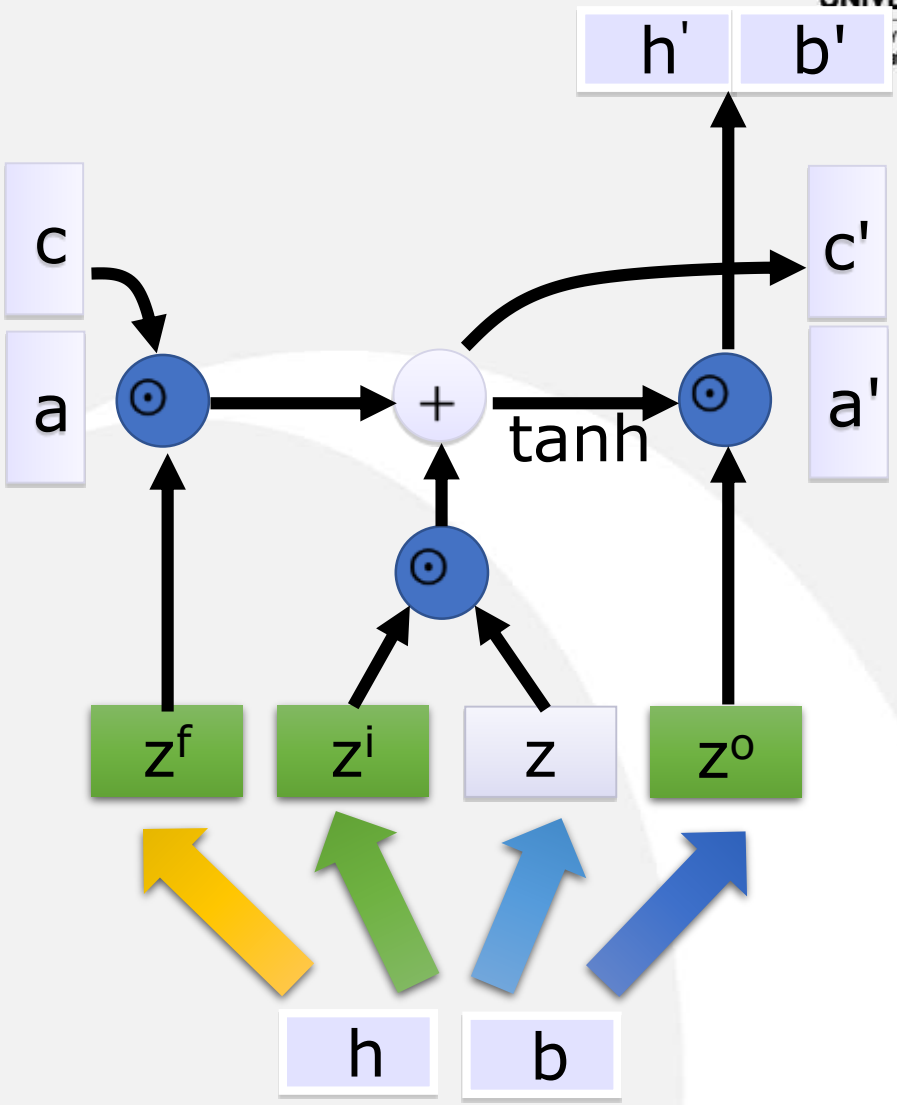
Grid LSTM

Memory for both ***time*** and ***depth***



Grid LSTM





You can generalize this to 3D, and more.