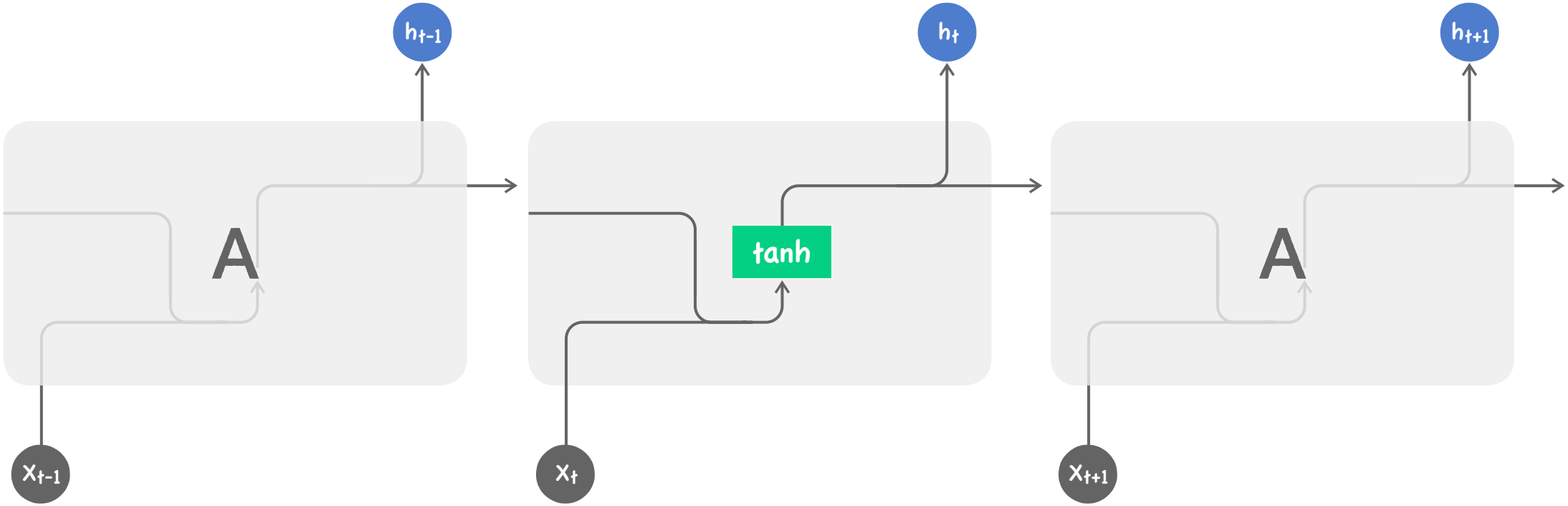
LECTURE 12-2

LSTM INTRODUCTION

Sung Kim <hunkim+ml@gmail.com>
http://hunkim.github.io/ml

NAVER | Clova

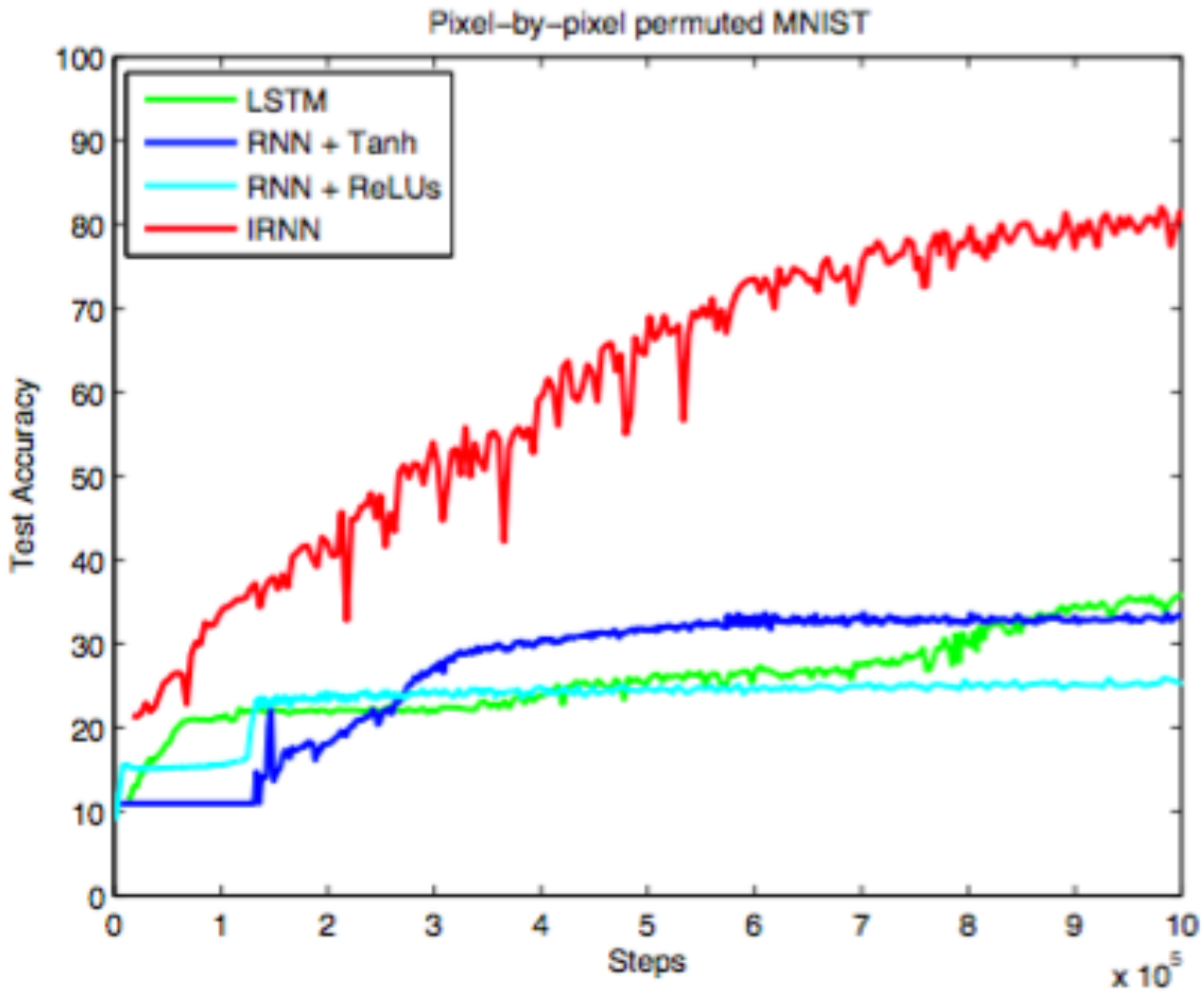


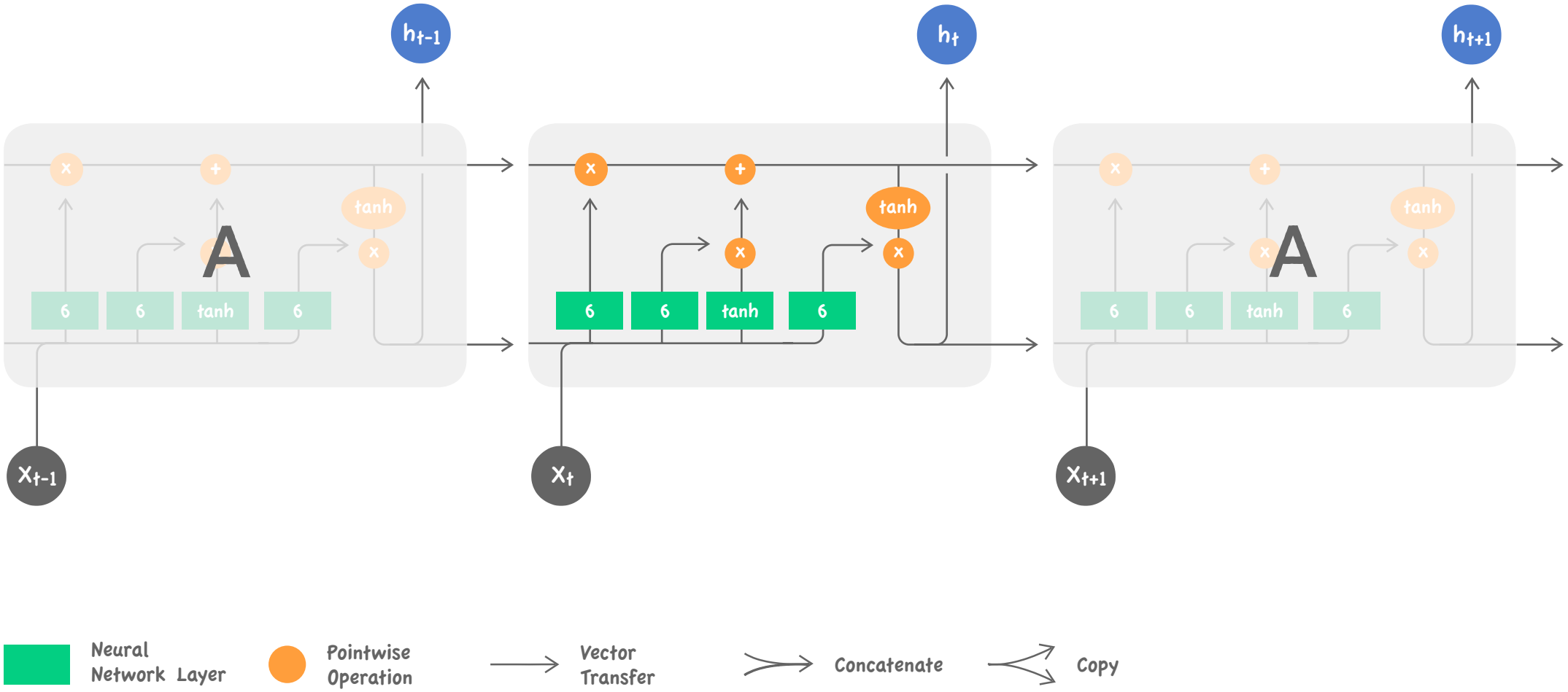
Given list of word vectors : $x_1, \dots, x_{t-1}, x_t, x_{t+1}, \dots, x_T$

At a single time step : $h_t = \sigma(W^{(hh)}h_{t-1} + W^{(hx)}x_{[t]})$

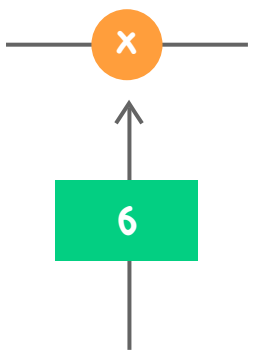
$\hat{y}_t = \text{softmax}(W^{(S)}h_t)$

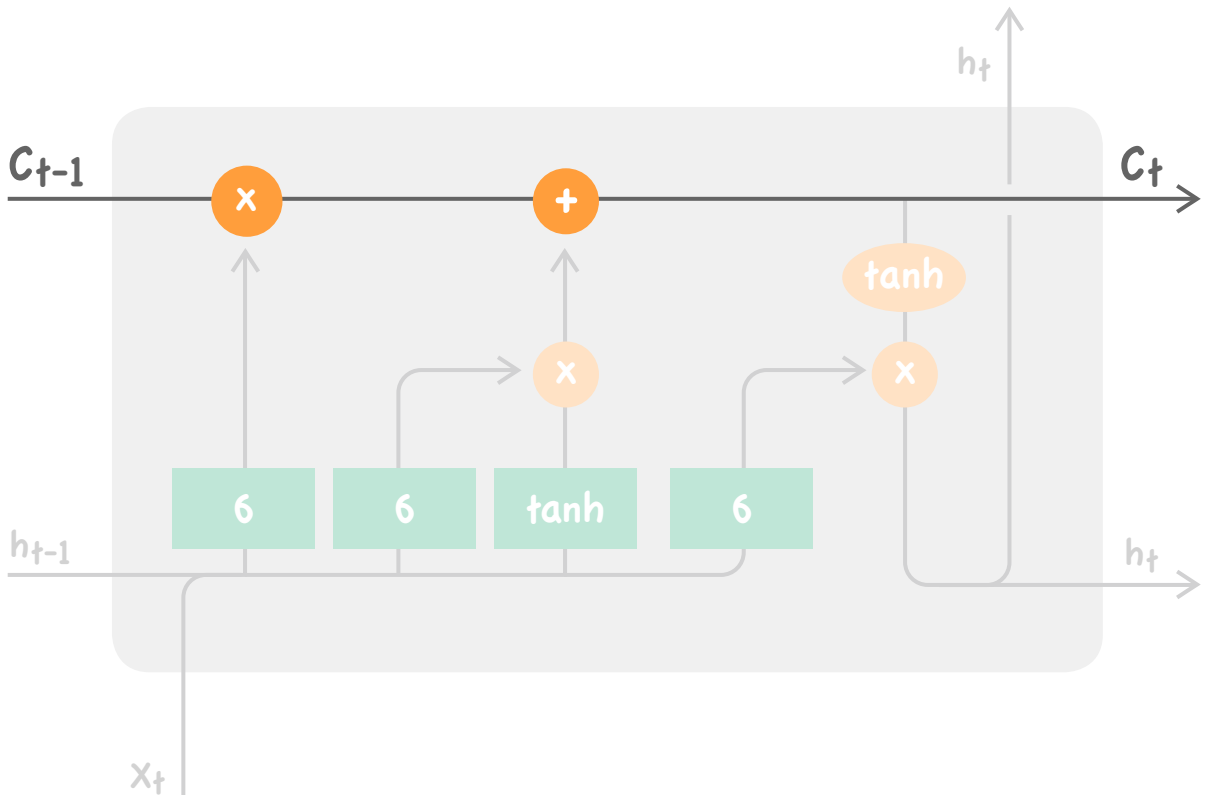
$\hat{P}(x_{t+1} = v_j \mid x_t, \dots, x_1) = \hat{y}_{t,j}$


Pixel-by-pixel permuted MNIST
LSTM
RNN + Tanh
RNN + ReLUs
IRNN
Test Accuracy
Steps
x 10^5
0
10
20
30
40
50
60
70
80
90
100
0
1
2
3
4
5
6
7
8
9
10

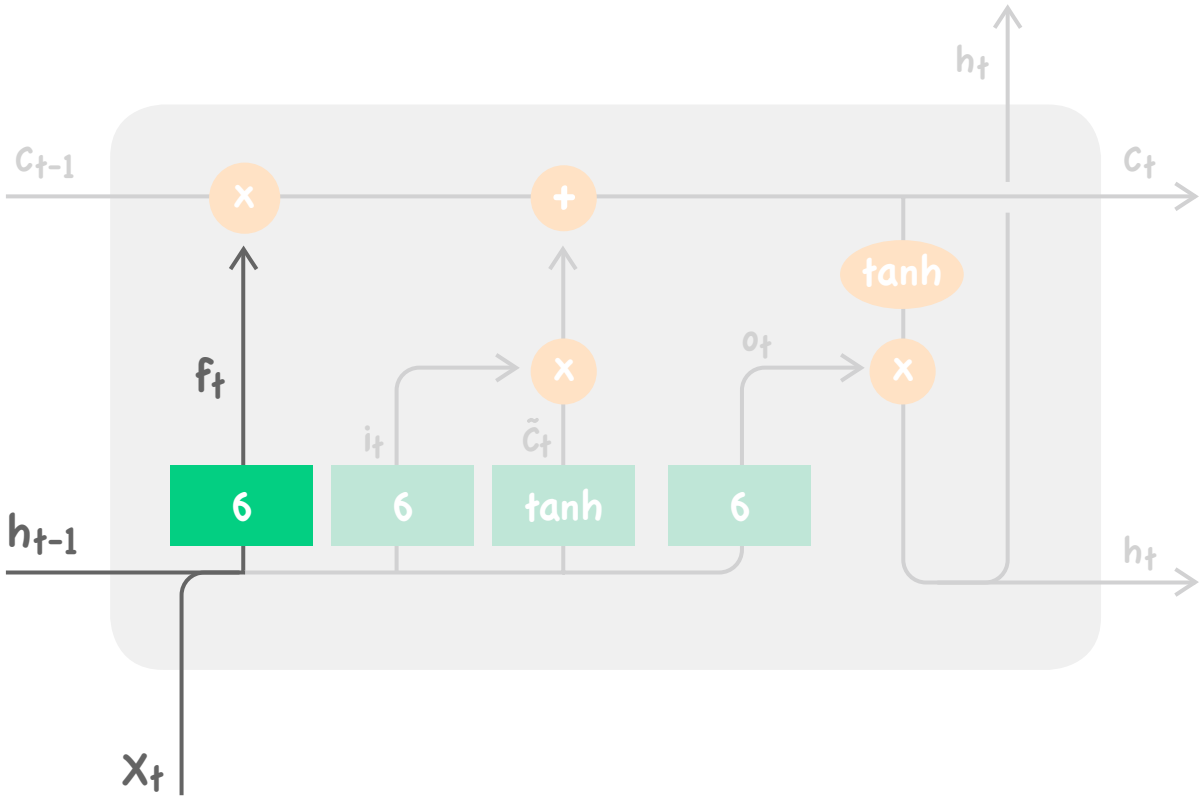

h_{t-1}
h_t
h_{t+1}
A
A
tanh
tanh
tanh
6
6
tanh
6
X_{t-1}
X_t
X_{t+1}
Neural Network Layer
Pointwise Operation
Vector Transfer
Concatenate
Copy

Gate

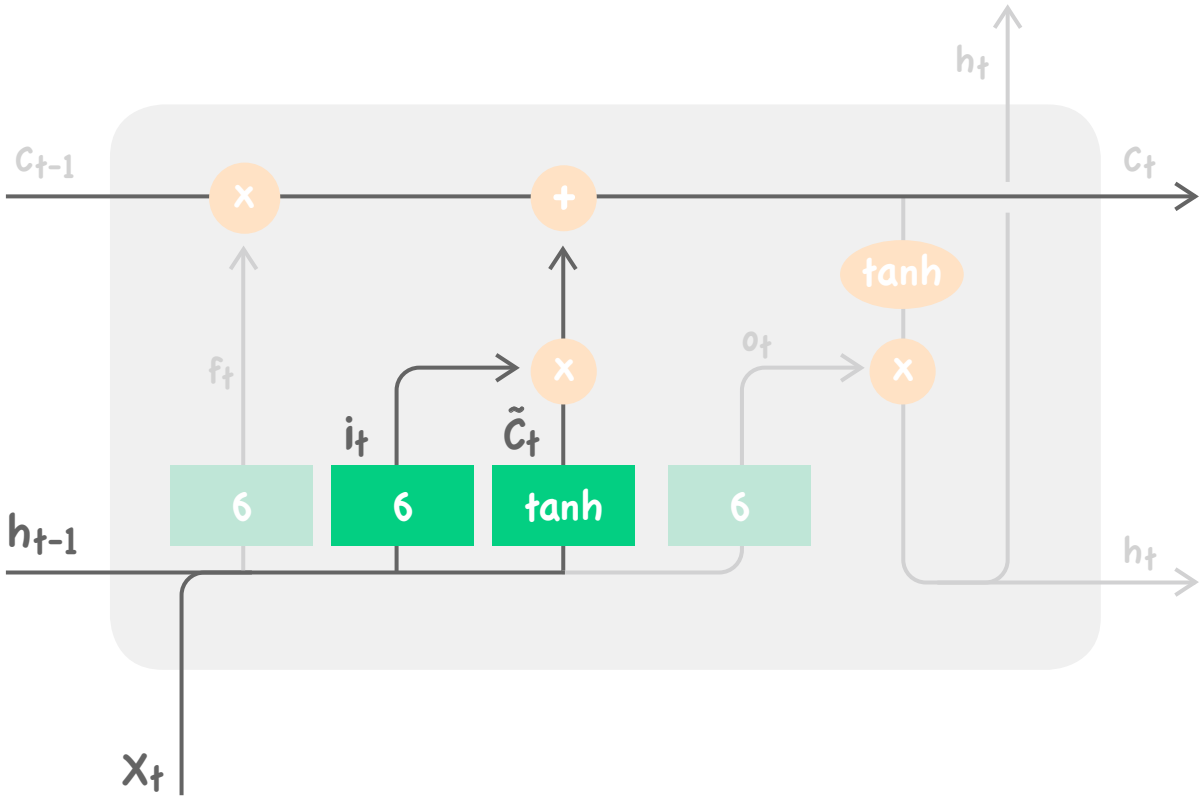




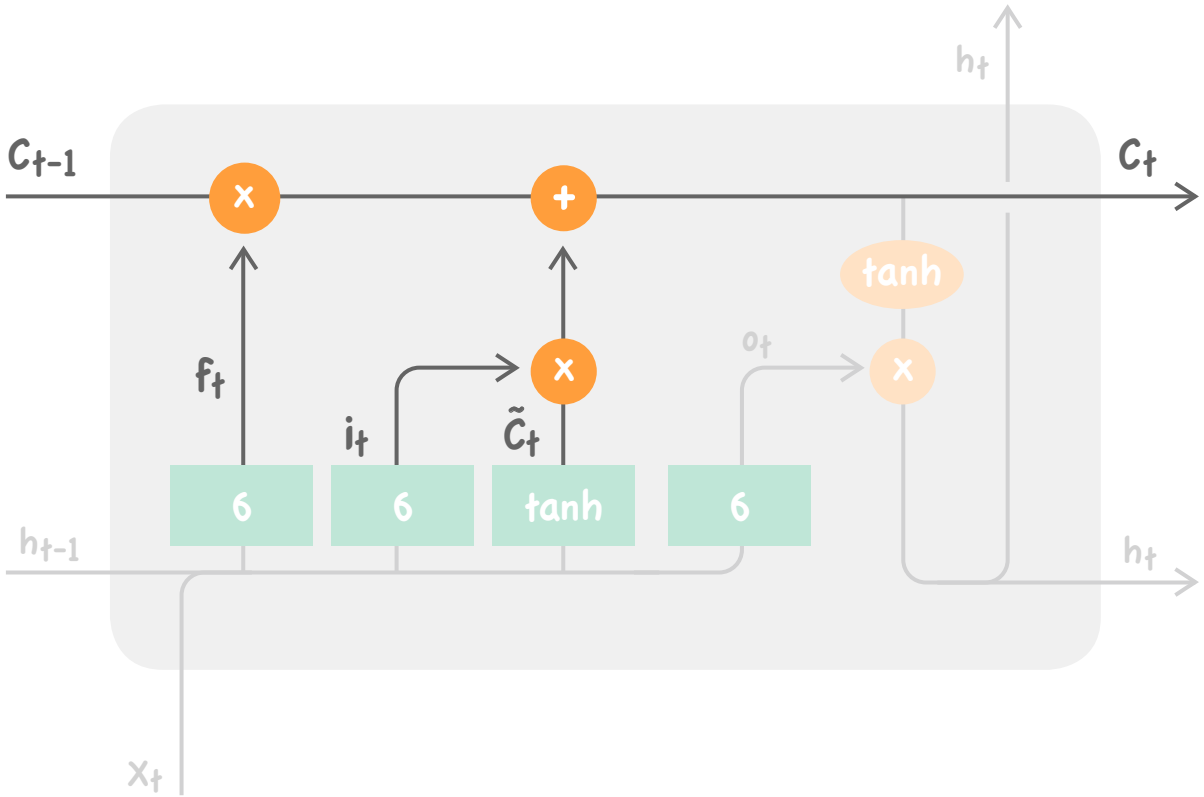
http://colah.github.io/posts/2015-08-Understanding-LSTMs/



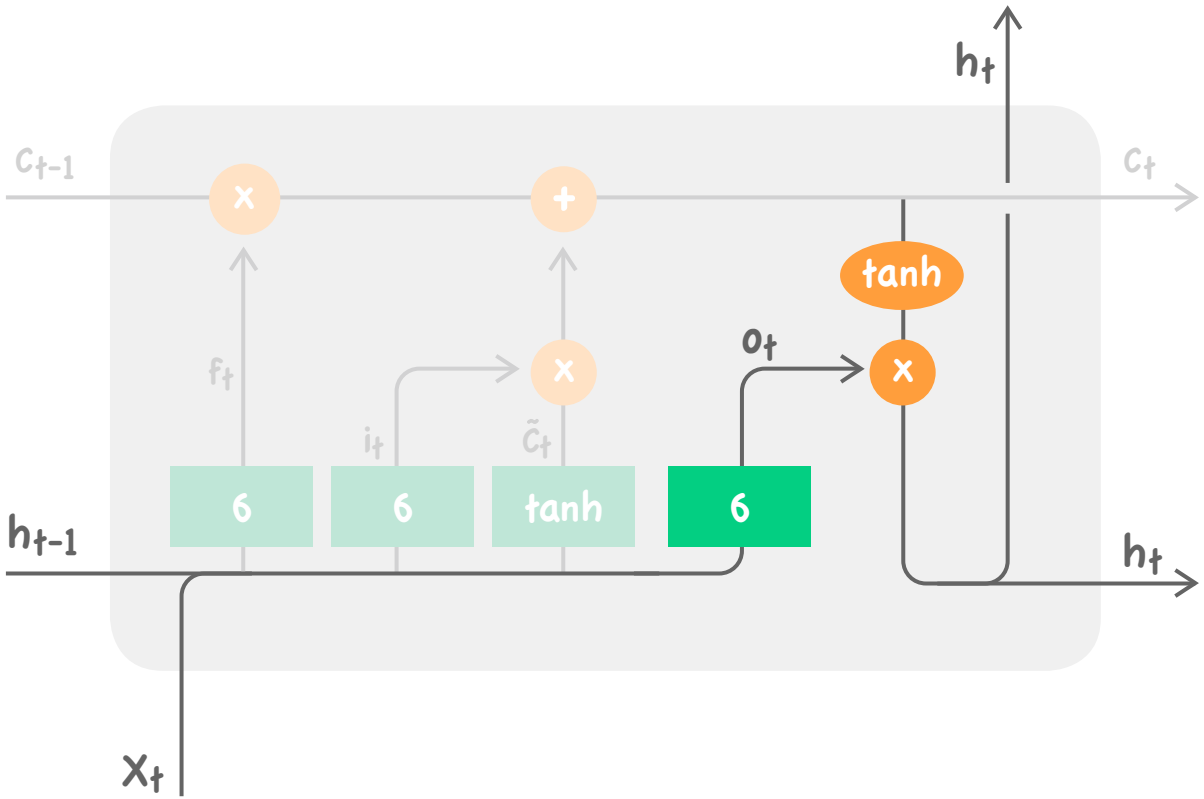
$$f_t = 6(W_f \cdot [h_{t-1}, X_t] + b_f)$$



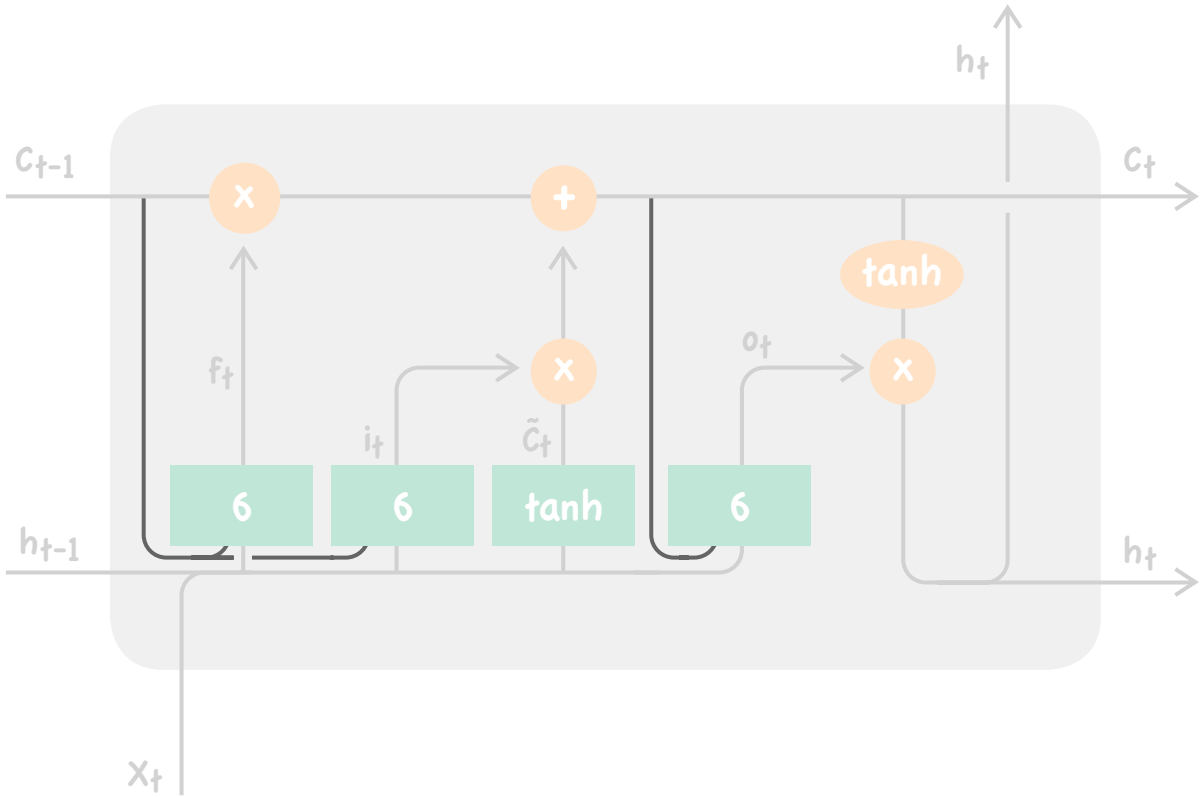
$$i_t = \sigma(W_i \cdot [h_{t-1}, X_t] + b_i)$$
$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, X_t] + b_c)$$



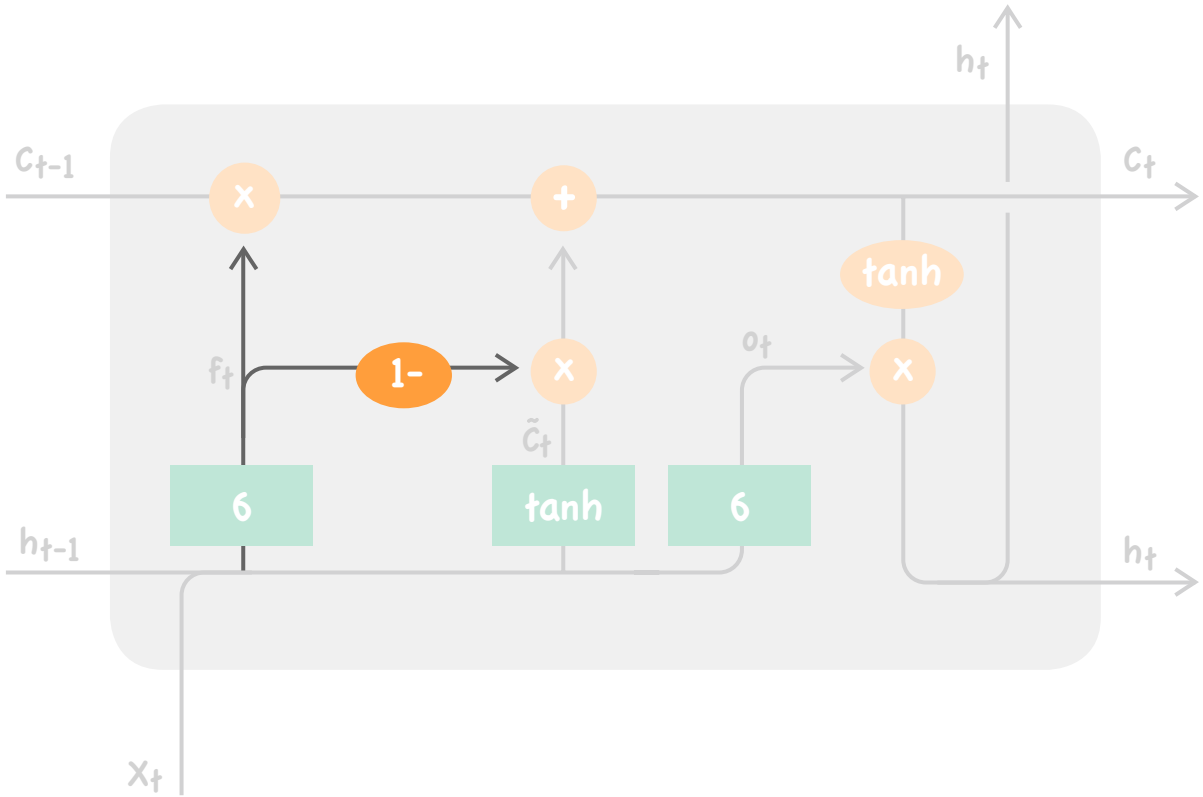
$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$



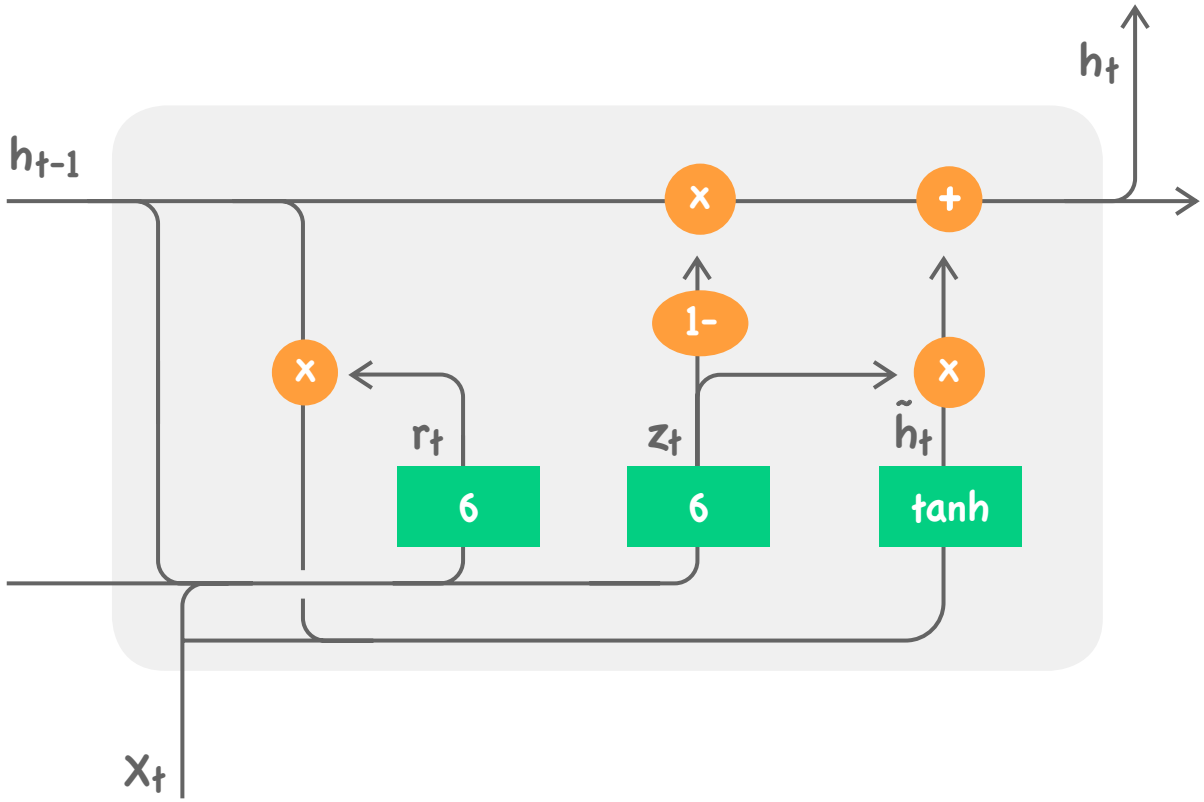
$$O_t = 6(W_O \cdot [h_{t-1}, X_t] + b_O)$$
$$h_t = O_t * \tanh(C_t)$$



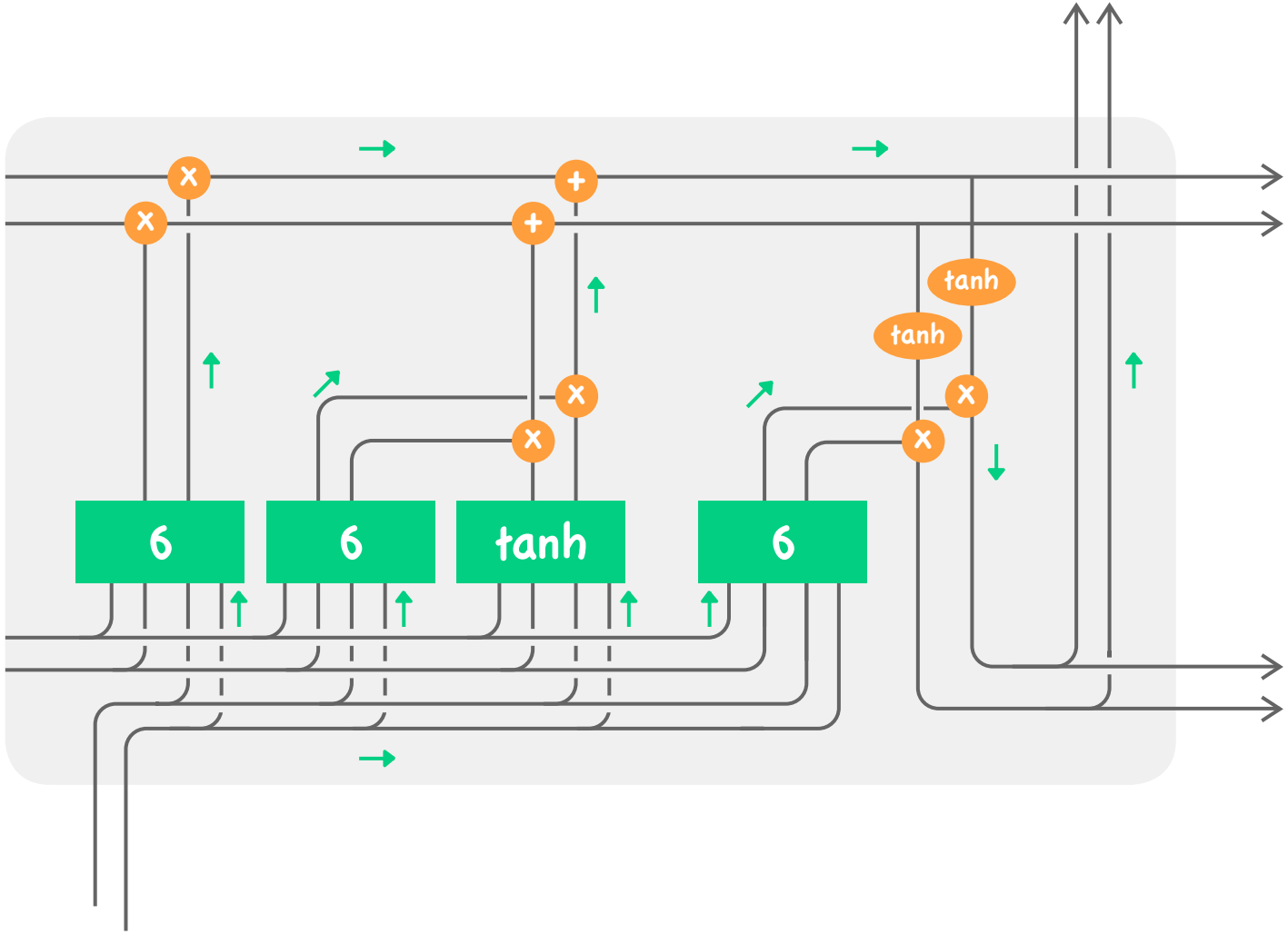
$$f_t = \sigma(W_f \cdot [C_{t-1}, h_{t-1}, X_t] + b_f)$$
$$i_t = \sigma(W_i \cdot [C_{t-1}, h_{t-1}, X_t] + b_i)$$
$$O_t = \sigma(W_o \cdot [C_{t-1}, h_{t-1}, X_t] + b_o)$$



$$C_t = f_t * C_{t-1} + (1 - f_t) * \tilde{C}_t$$



$$z_t = 6(W_z \cdot [h_{t-1}, X_t])$$
$$r_t = 6(W_r \cdot [h_{t-1}, X_t])$$
$$\tilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, X_t])$$
$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$


tanh
tanh
6
6
tanh
6

NEXT LECTURE

RNN / LSTM CASE STUDY