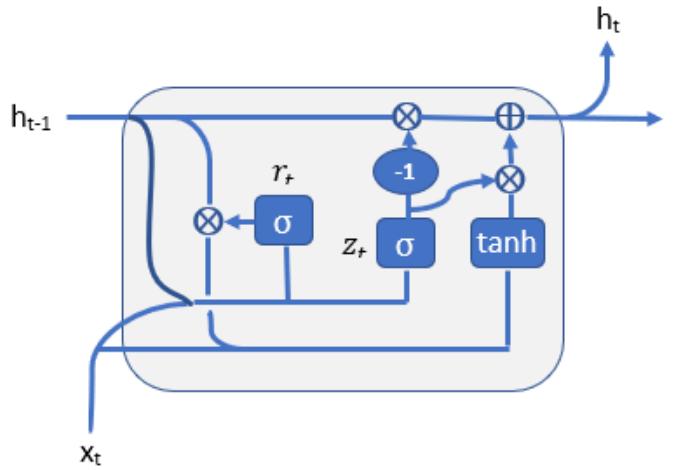


Gated Recurrent Unit (GRU)

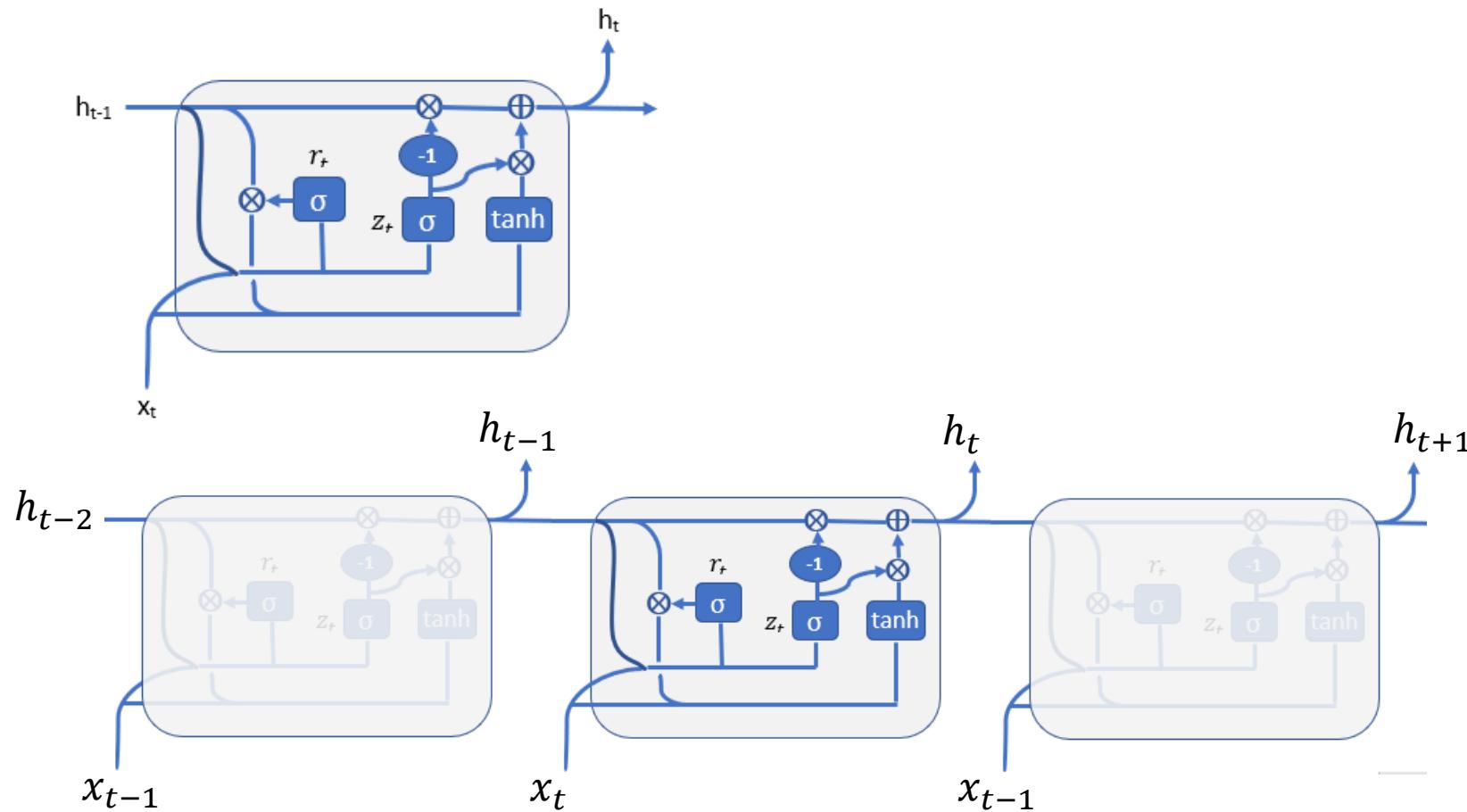
Unlike LSTM, GRU has only one State and only two Gates

Hidden State
&
Update Gate and Reset Gate

Gated Recurrent Unit (GRU)

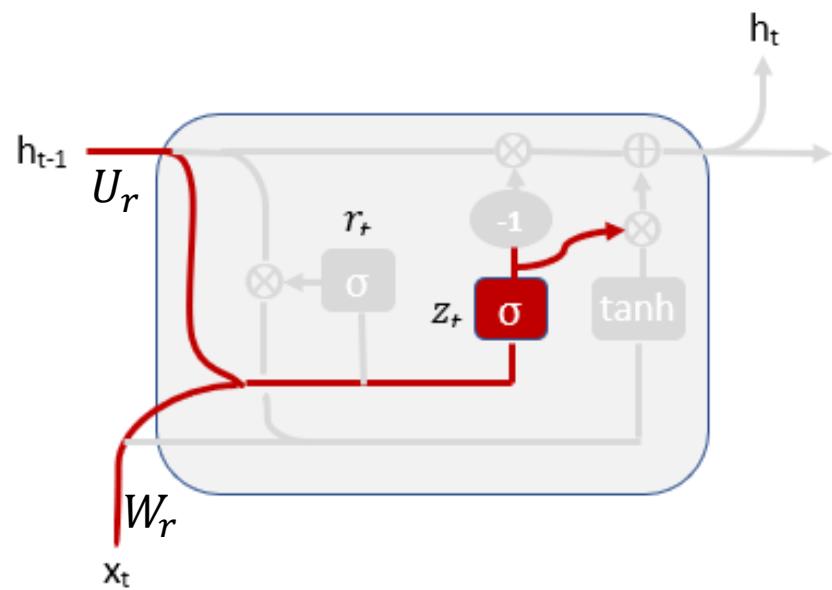


Gated Recurrent Unit (GRU)



Update Gate

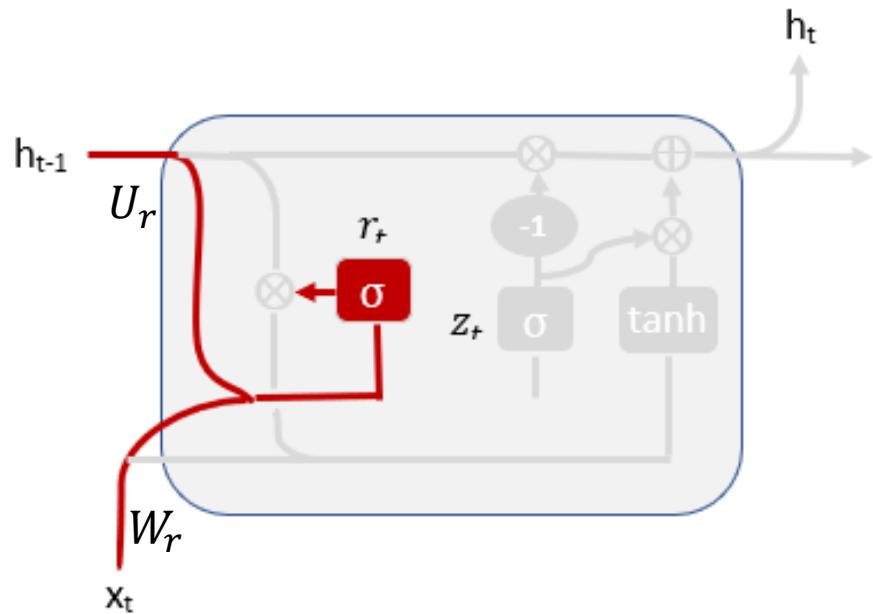
How much of the past information need to be considered to the next time step.



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

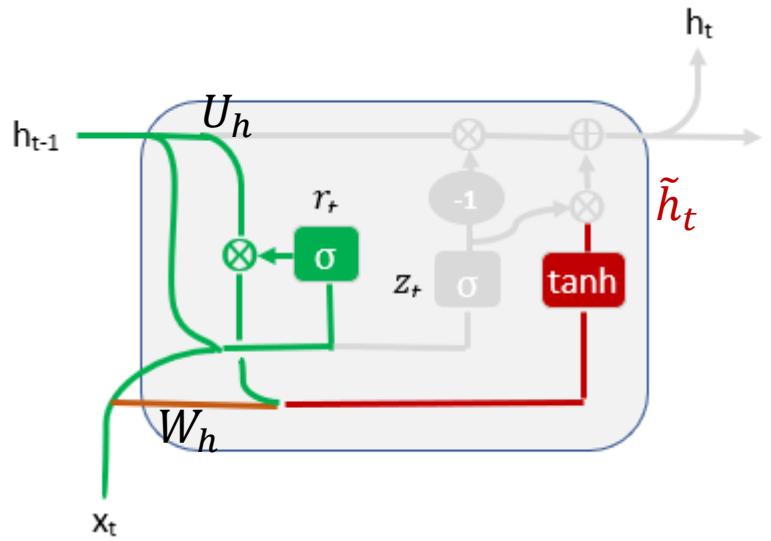
Reset Gate (r_t)

How much of the past information to be forgotten?



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

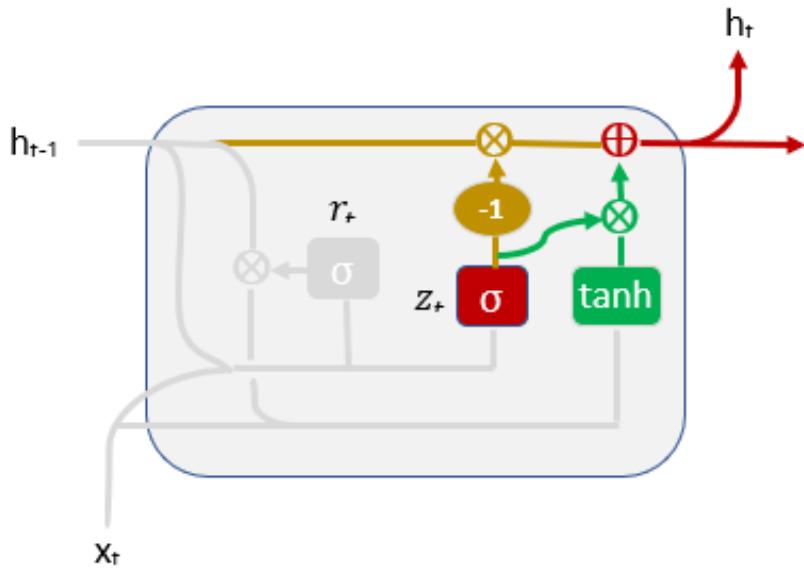
Internal Information (Current Memory Content)



$$\tilde{h}_t = \tanh(W_h \cdot x_t + r_t \otimes U_h \cdot h_{t-1})$$

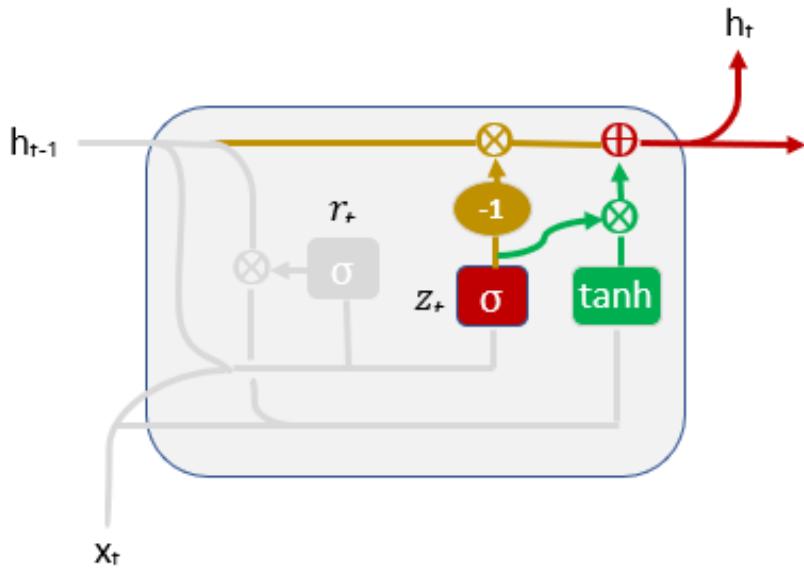
Consider the present input and useful information of the previous hidden state to produce new internal information

New Hidden State (h_t)



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

New Hidden State (h_t)

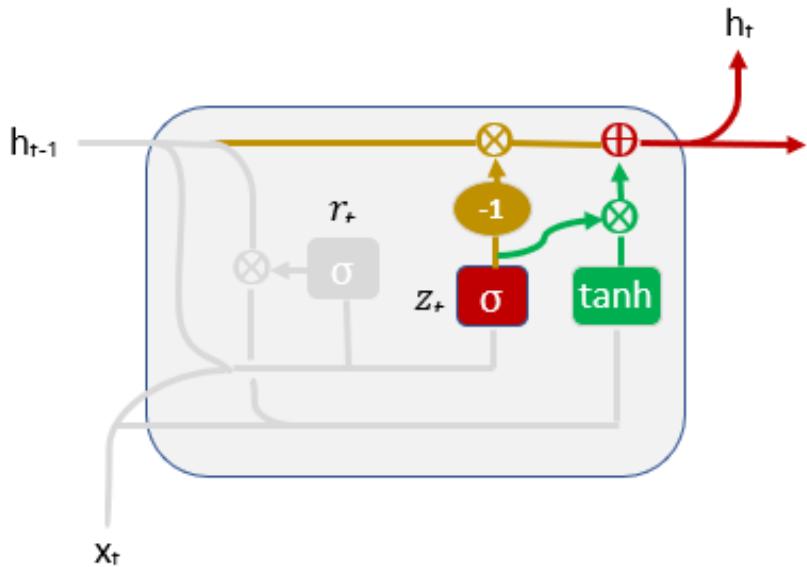


$$h_t = (1 - z_t) \otimes h_{t-1} \oplus z_t \otimes \tilde{h}_t$$

Previous hidden state at
the position which are
ignored in internal
information

Useful information from
the internal information

New Hidden State (h_t)



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

A hand-drawn example shows two vectors being multiplied and added. The first vector is $[1, 0, 0.3, 1, 0]$ and the second is $[b, g, h, i, j]$. They are multiplied element-wise (indicated by a circled \otimes) to produce $[b, g, 0.3h, i, j]$. This result is then added to another vector $[0, 1, 0.7, 0, 1]$ (indicated by a circled $+$) to produce the final result $[a, b, c, d, e]$.

Summary

One State

- Hidden State $h_t = (1 - z_t) \otimes h_{t-1} \oplus z_t \otimes \tilde{h}_t$

Two Gates

- Update Gate $Z_t = \sigma(W_z \cdot x_t + U_z \cdot h_{t-1})$
- Reset Gate $r_t = \sigma(W_r \cdot x_t + U_r \cdot h_{t-1})$