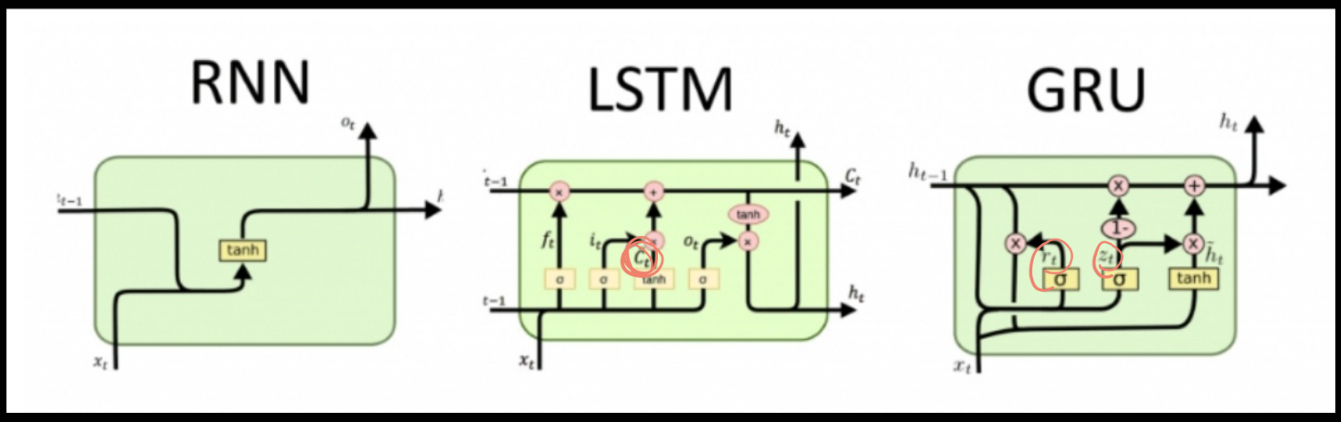
Gated Recurrent Unit (GRU)

GRUs simplify the LSTM architecture by combining the input and forget gates into a single "update gate" and also merging the cell state and hidden state. This results in a more efficient model that can perform similarly to LSTMs on many tasks but with fewer parameters and computational complexity.

- **Reset Gate**: This gate decides how much of the previous hidden state will influence the new candidate hidden state. If the reset gate is close to 0, it effectively makes the candidate state ignore the previous state, allowing the model to drop any irrelevant information from the past. This can be helpful when the previous state is not useful for the current prediction.

- **Update Gate**: This gate determines how much of the previous hidden state should be carried over to the next state, and how much should be updated with new information from the current input. It's a blend of the old and new information, preserving long-term dependencies while adding necessary short-term details. The update gate helps in balancing between the information from the past and the new information from the input.



Update Gate

$$z_t = \sigma(W^{(z)}x_t + U^{(z)}h_{t-1})$$

Reset Gate

$$r_t = \sigma(W^{(r)}x_t + U^{(r)}h_{t-1})$$

Current Memory Content

$$h_t' = \tanh(Wx_t + r_t \odot Uh_{t-1})$$