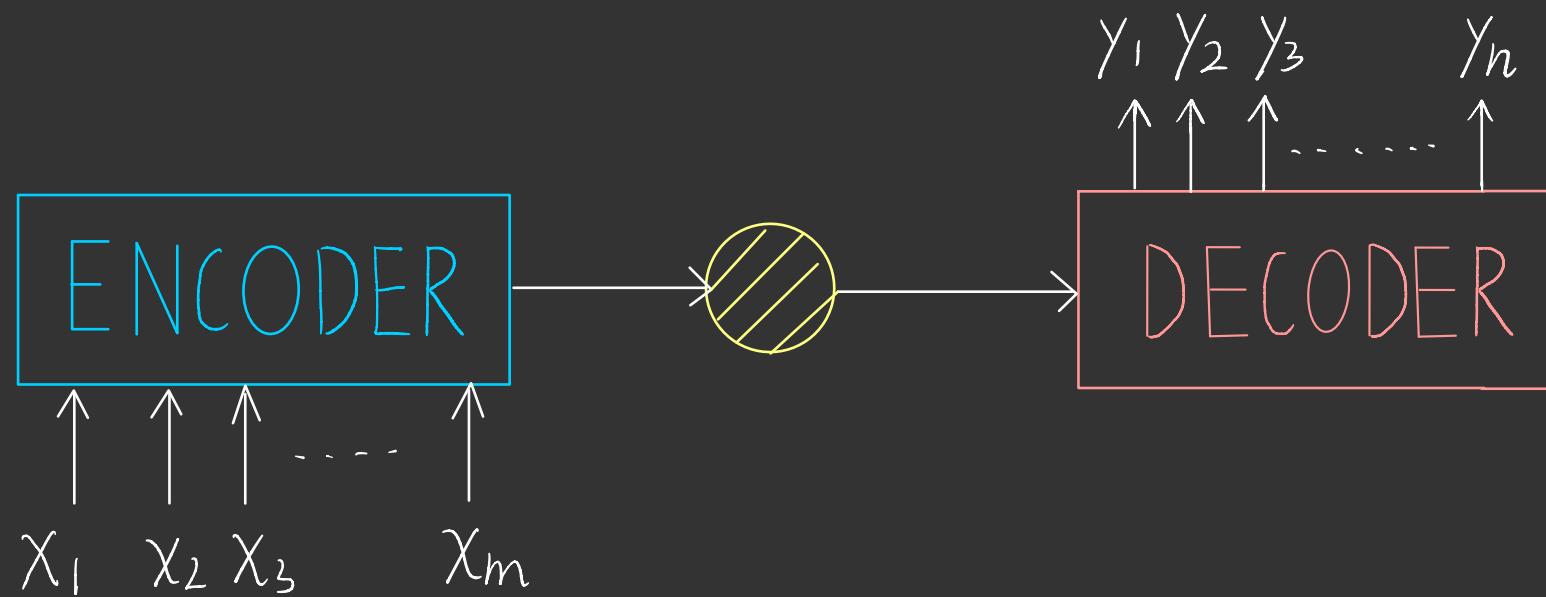
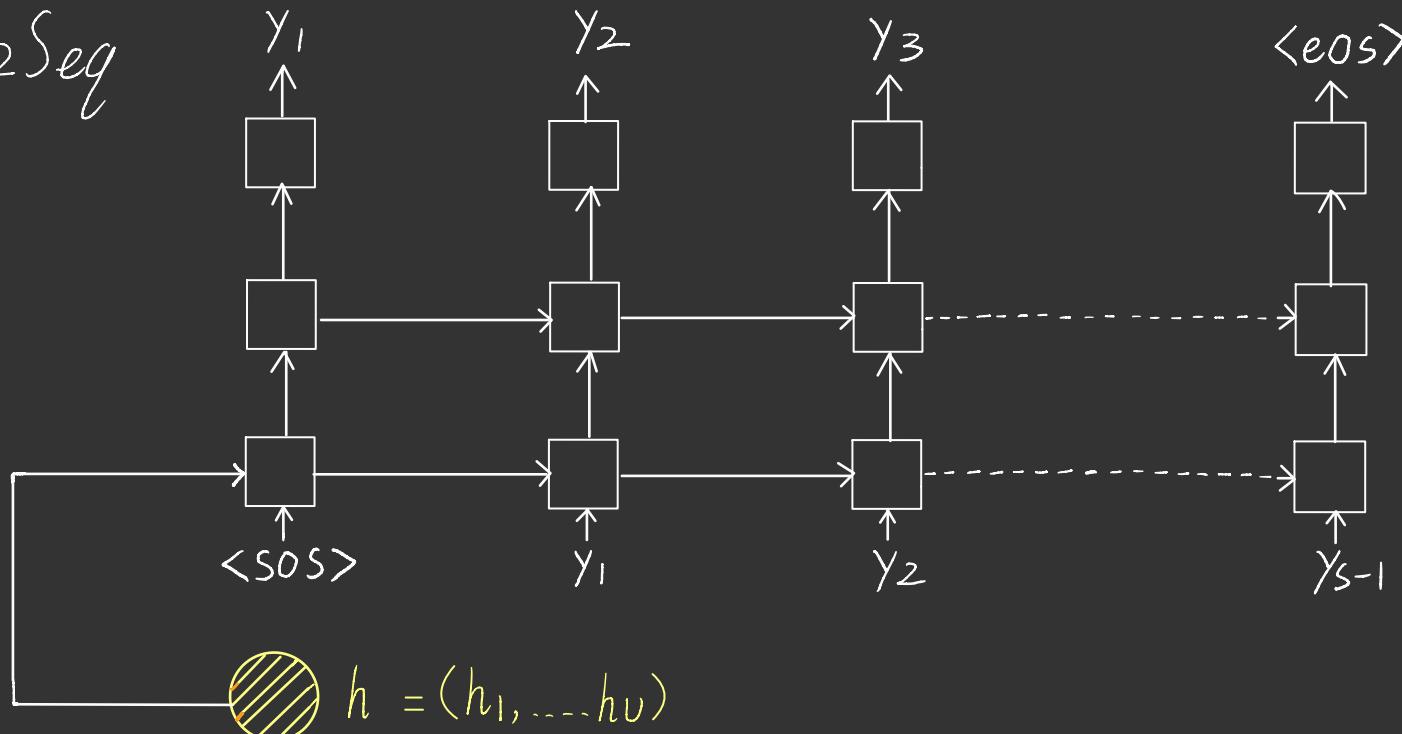


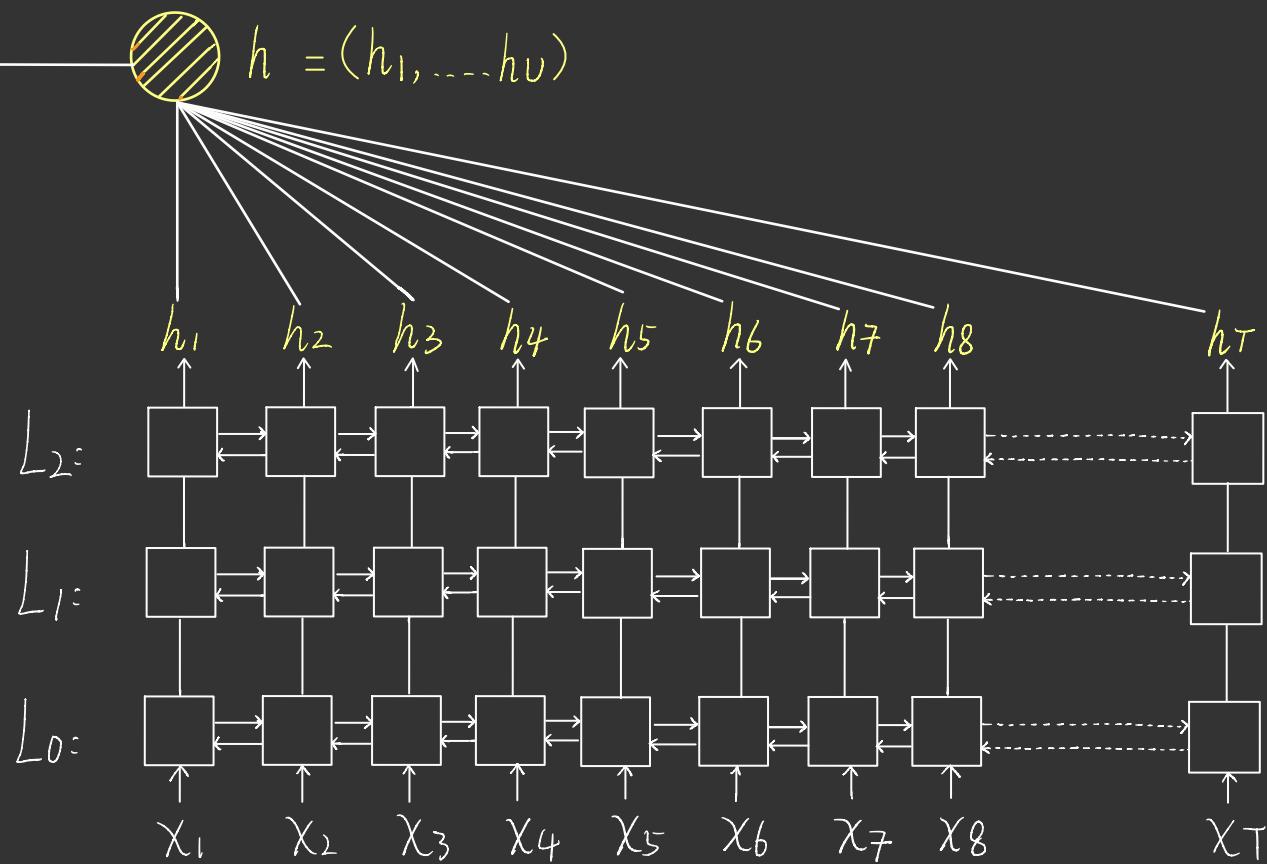
Seq2Seq

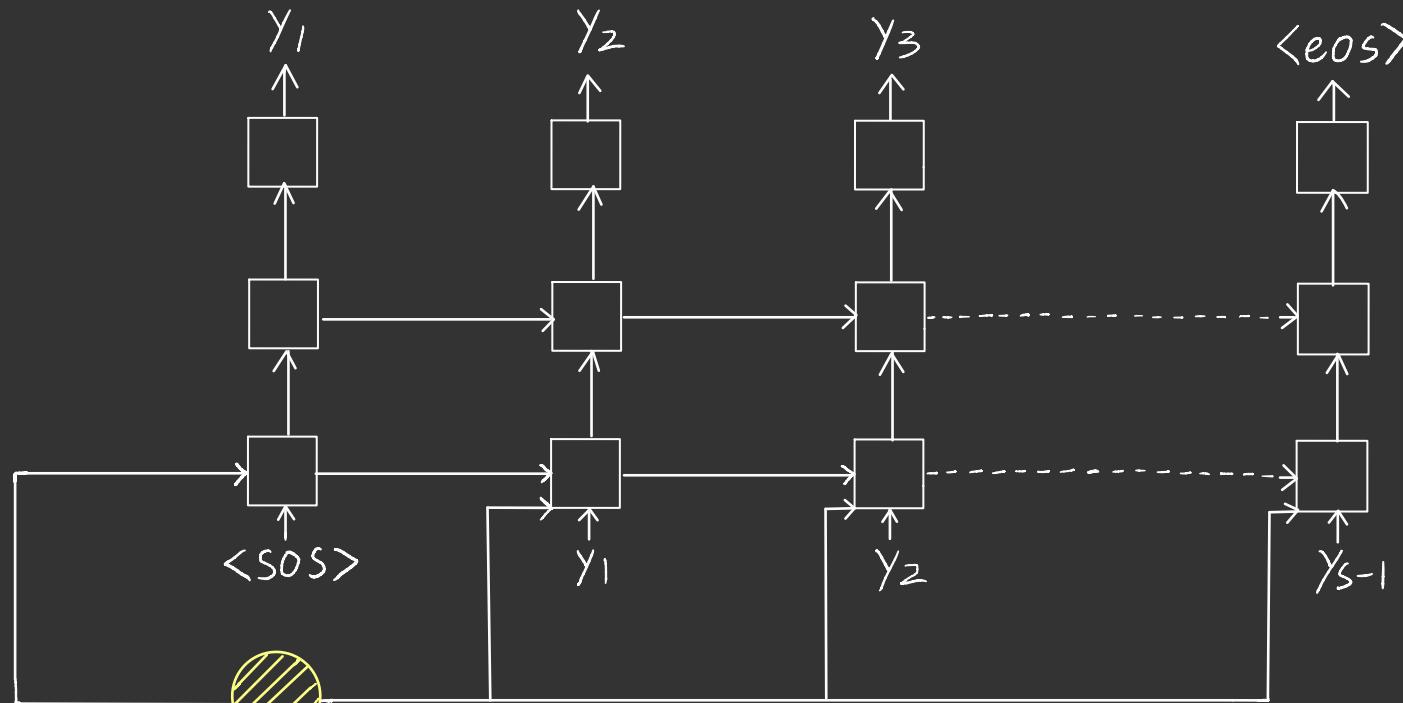


Vanilla Seq2Seq

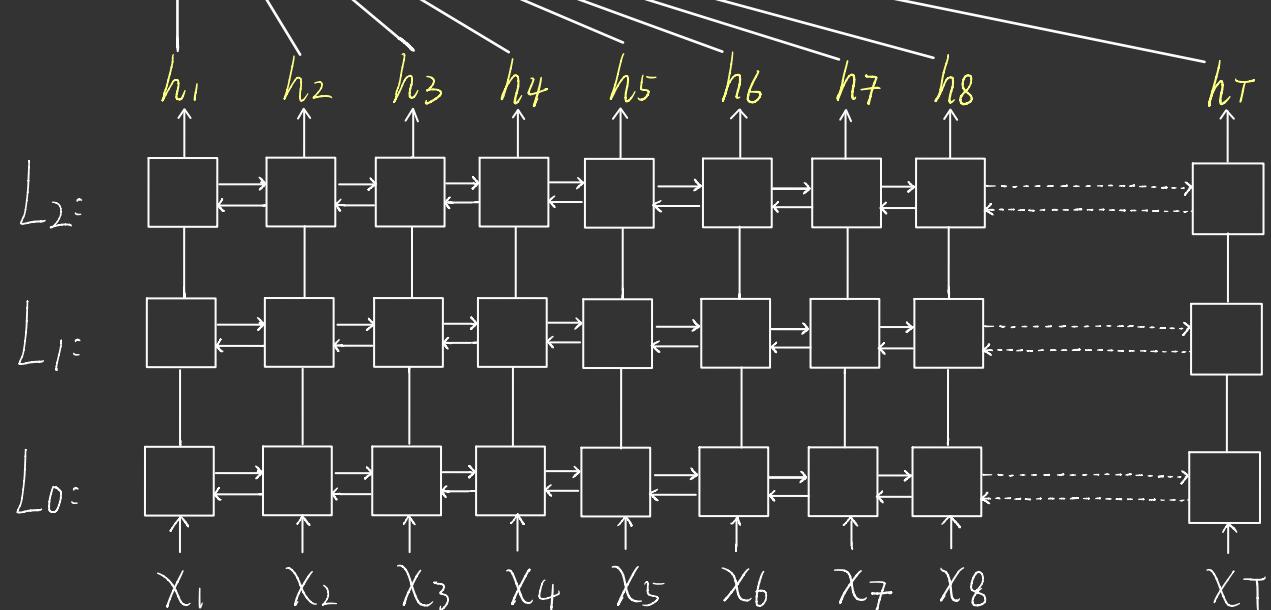


$$h = (h_1, \dots, h_T)$$

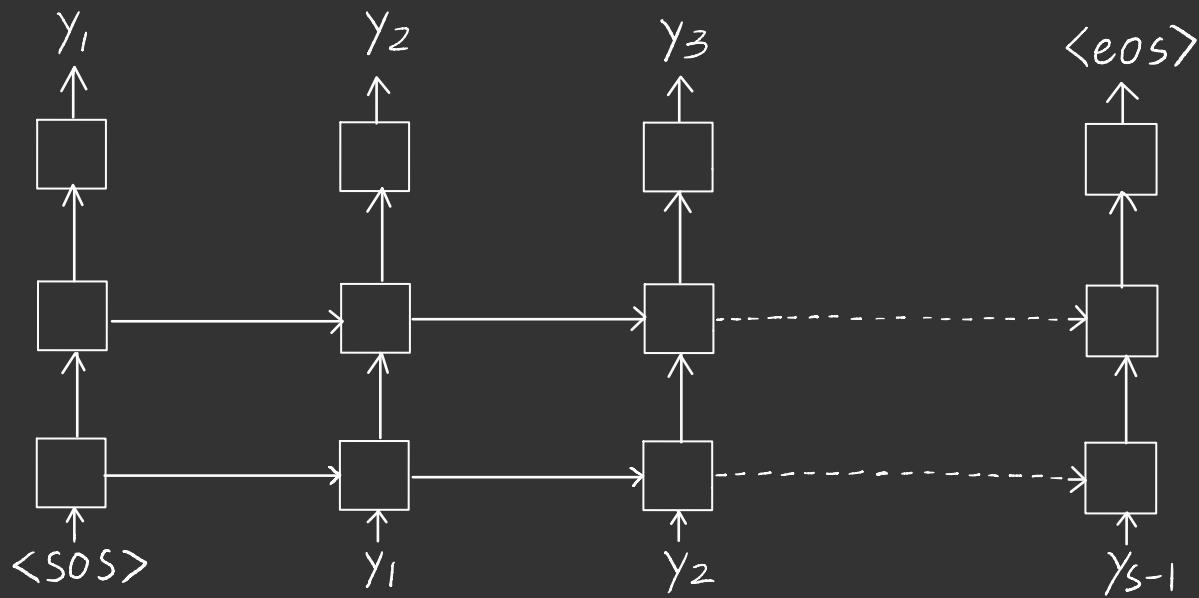




$$h = (h_1, \dots, h_U)$$



Attention

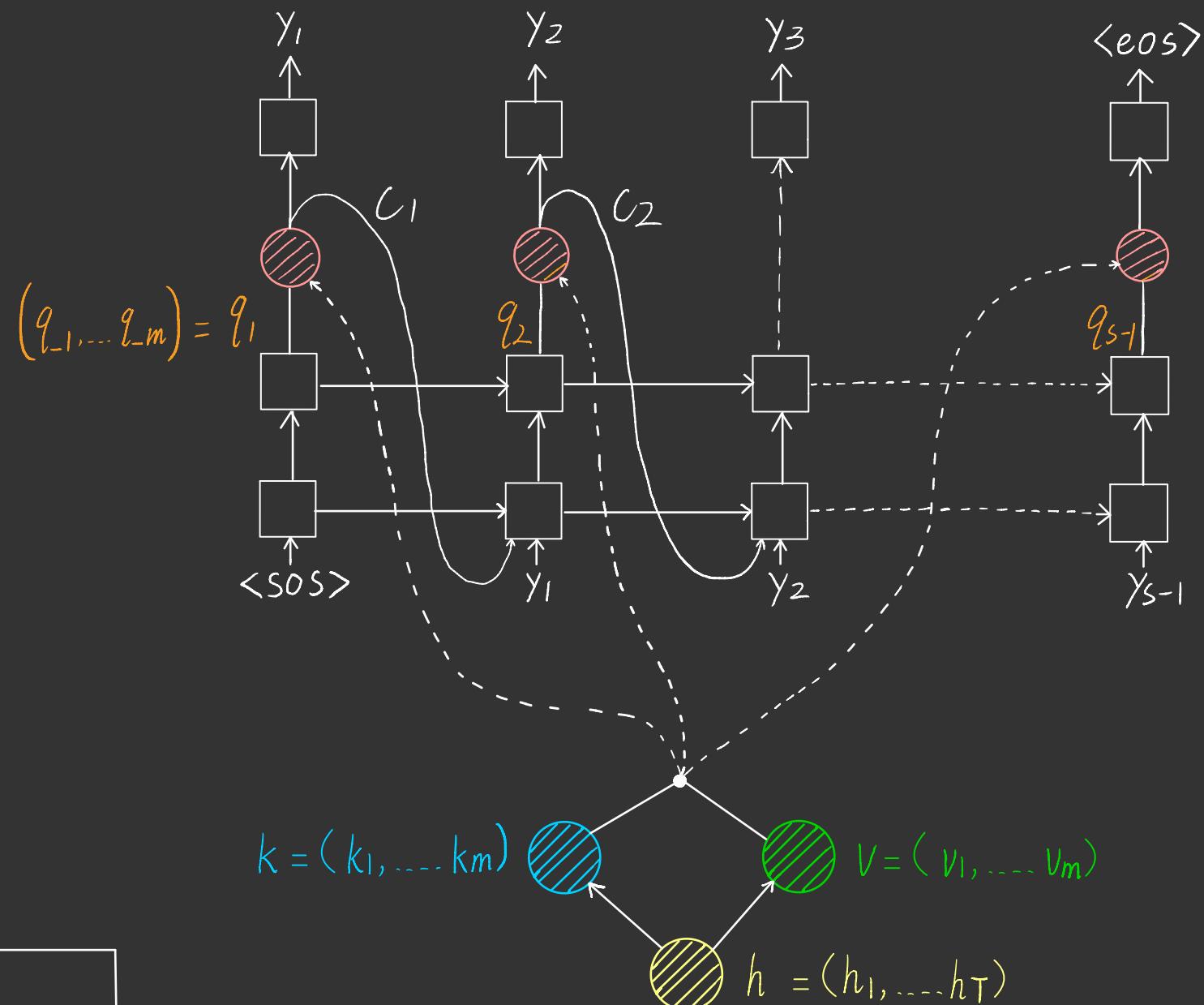


$$\text{h} = (h_1, \dots, h_T)$$

h_1	h_2	h_3	h_4	h_5	\dots	h_T
-------	-------	-------	-------	-------	---------	-------

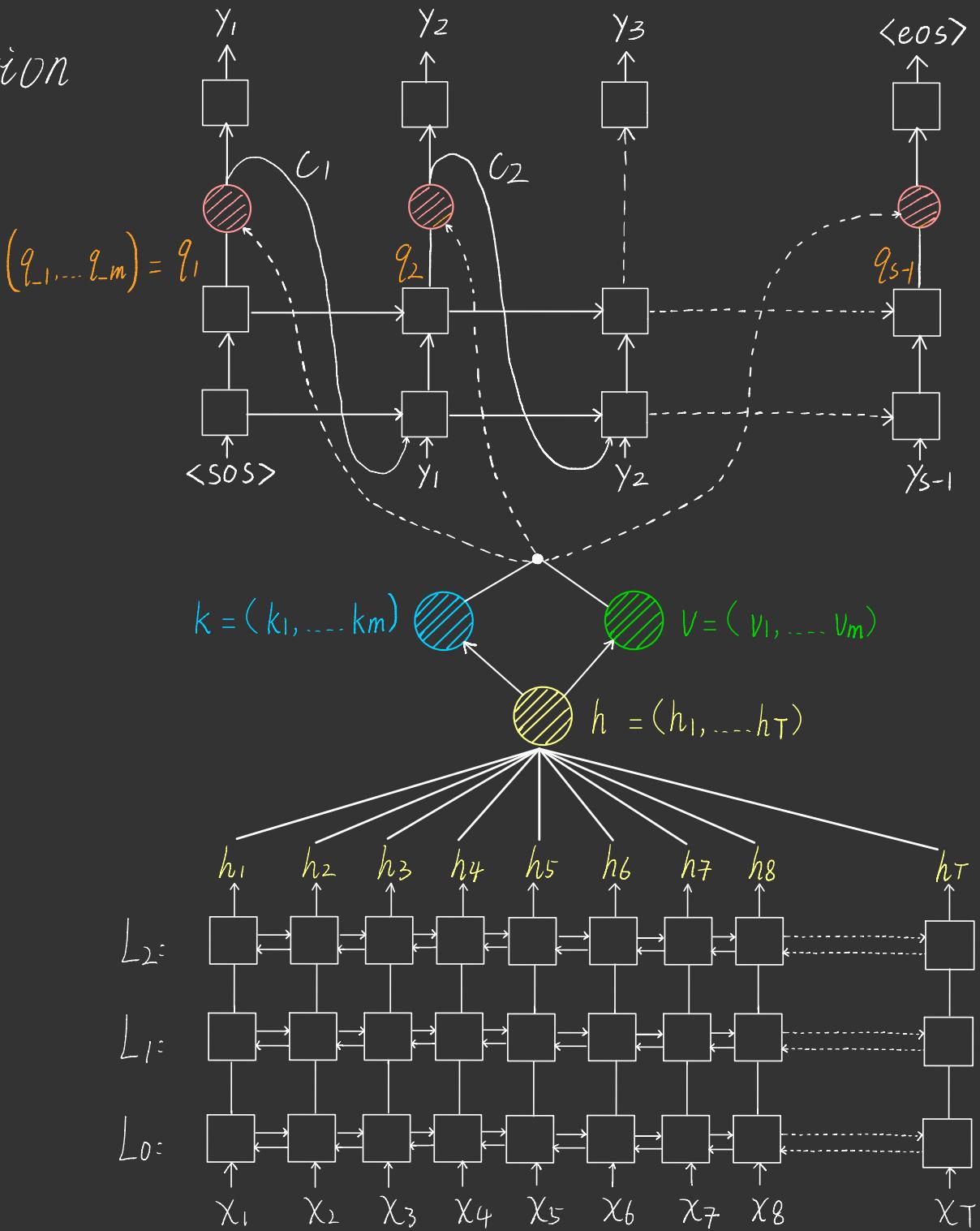
weight: $\alpha_1 \ \alpha_2 \ \alpha_3 \ \alpha_4 \ \alpha_5 \ \dots \ \alpha_T$

Eg.

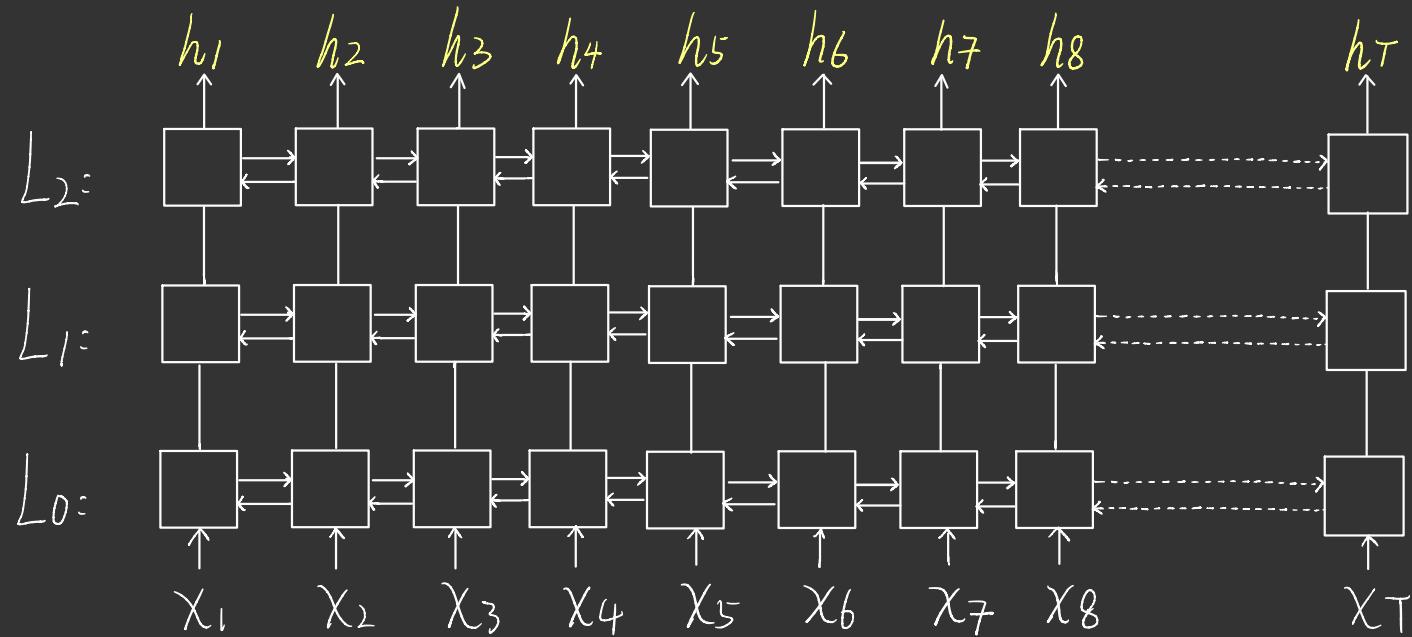


$C = \text{cos-sim}(q, k) \cdot V$

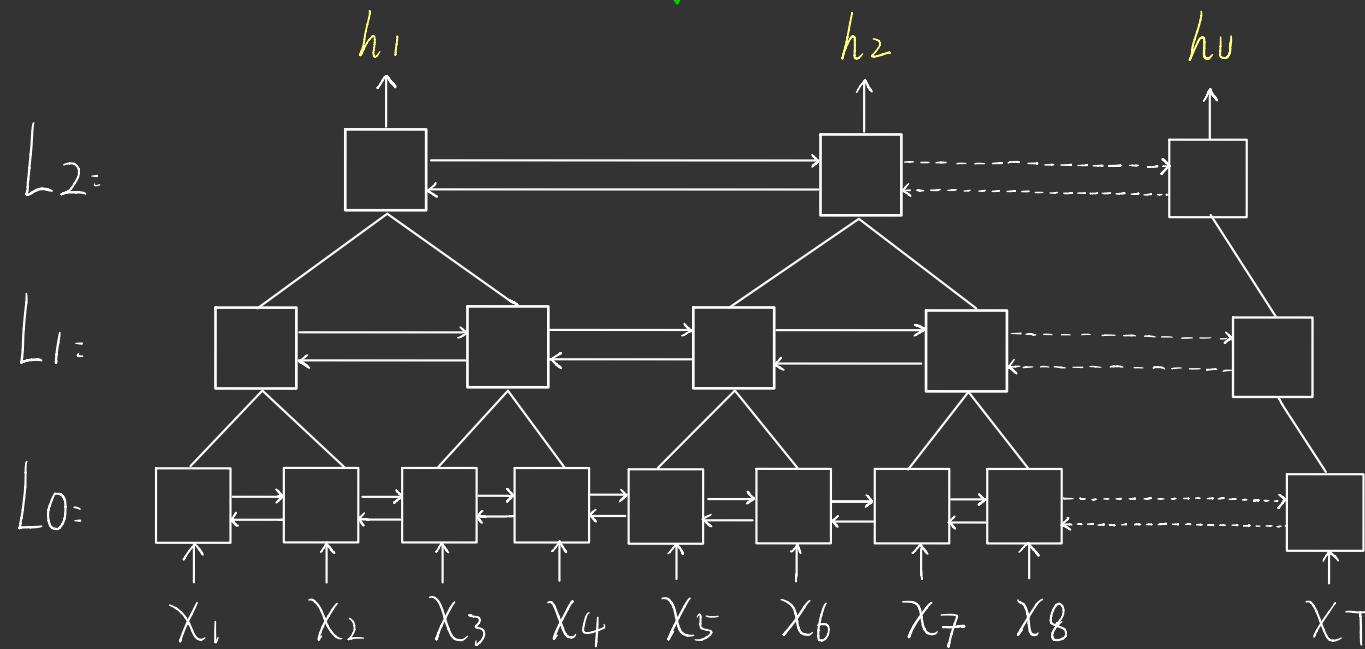
Seq2Seq + Attention



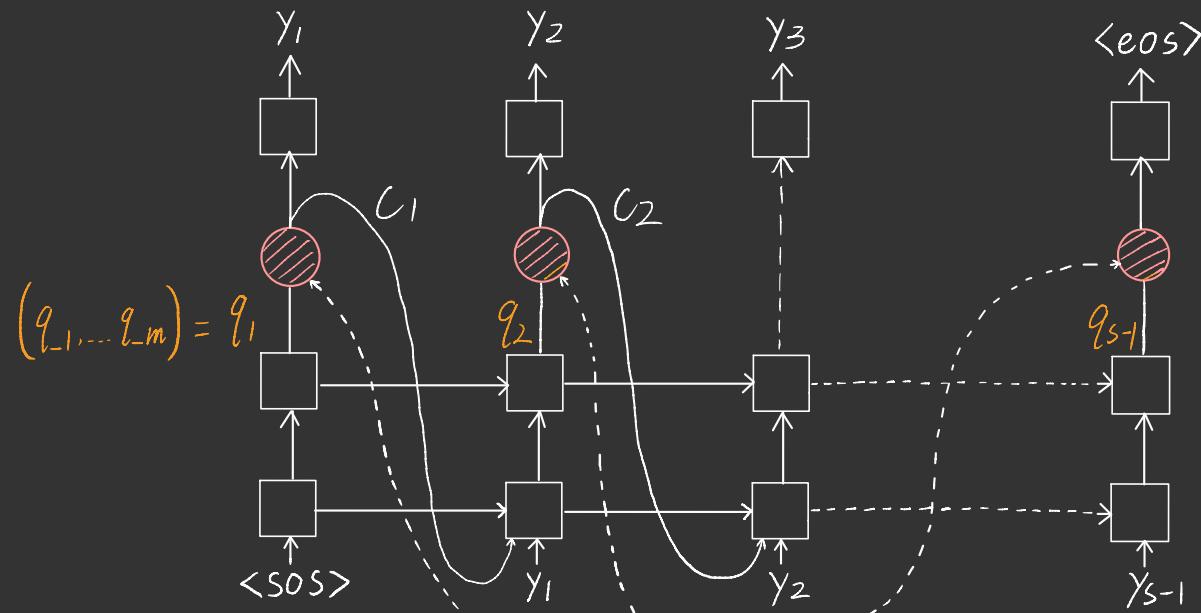
BLSTM



pBLSTM



LAS

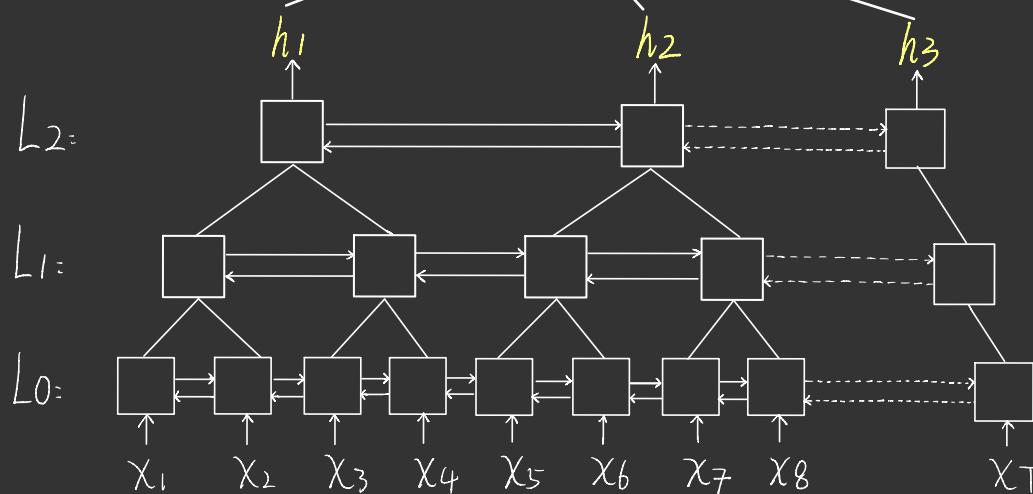


$$(q_1, \dots, q_m) = q$$

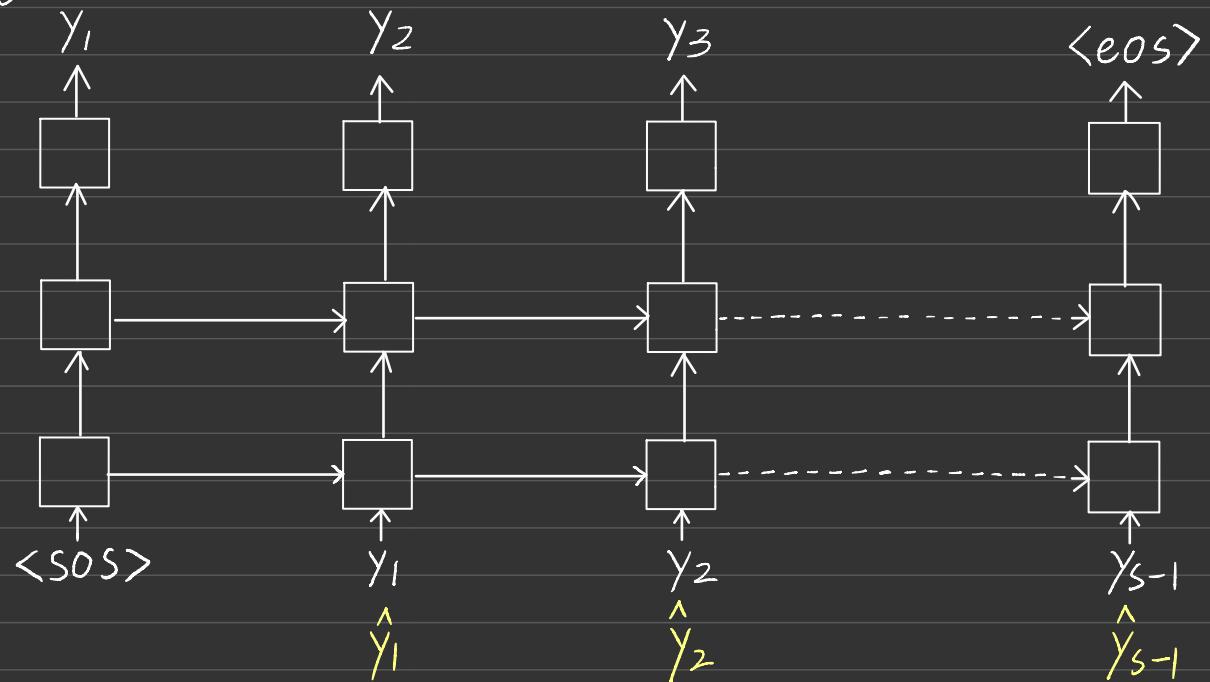
$$k = (k_1, \dots, k_m)$$

$$h = (h_1, \dots, h_T)$$

$$V = (v_1, \dots, v_m)$$



Teacher forcing:



Gumble noise

Data Augmentation

