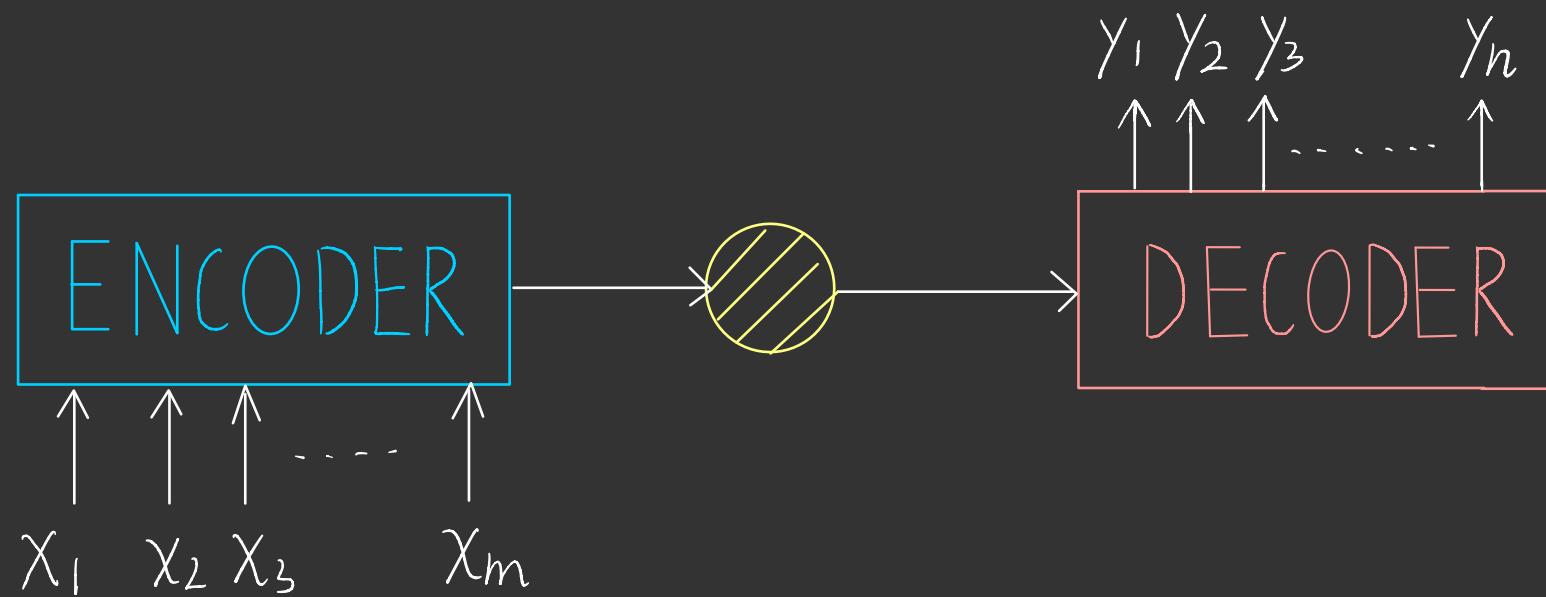
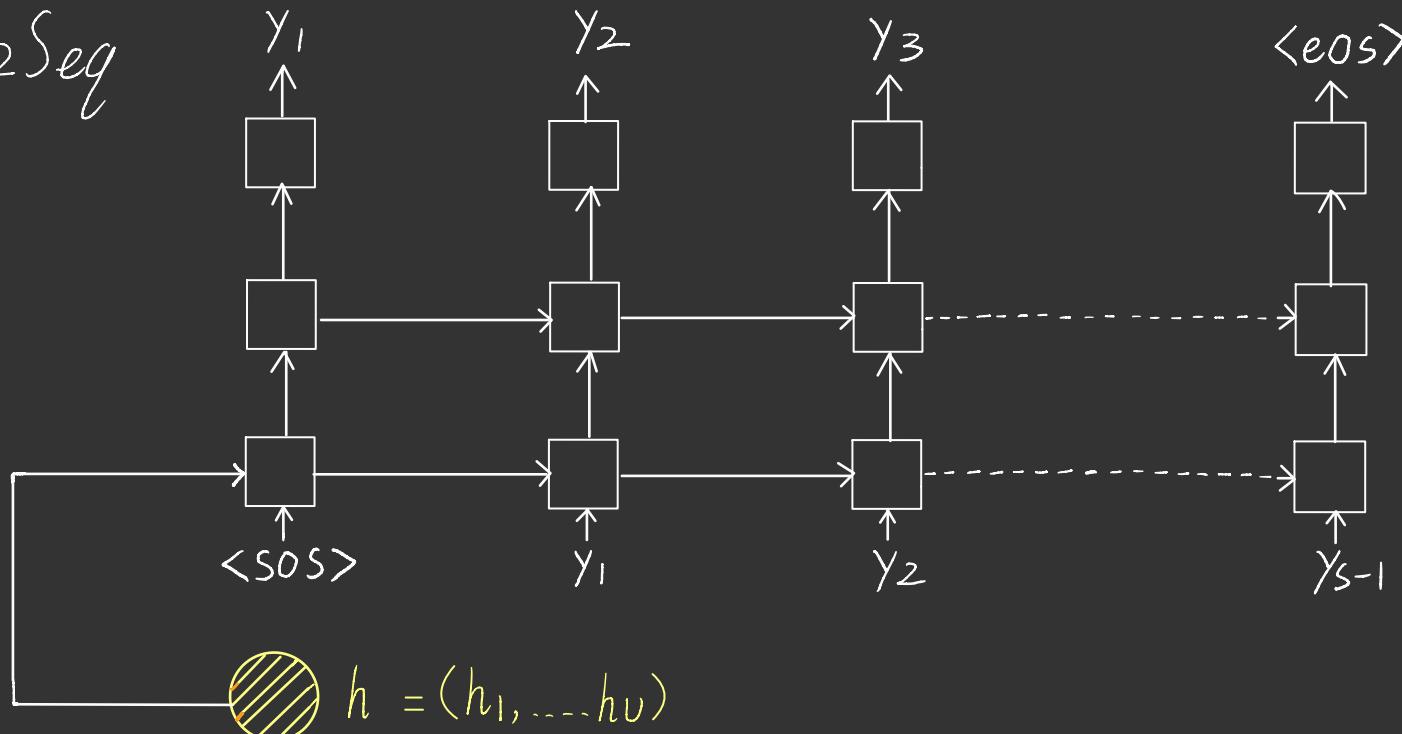


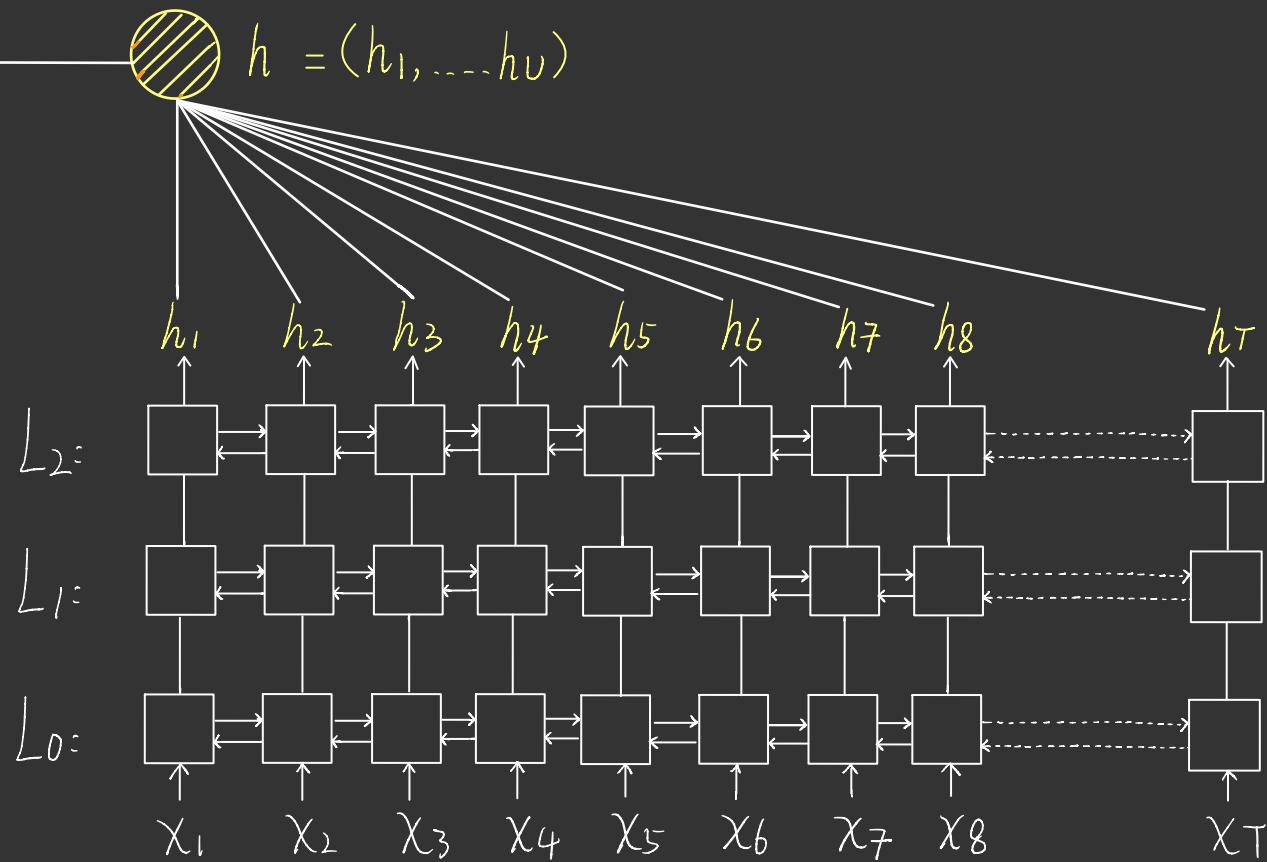
Seq2Seq

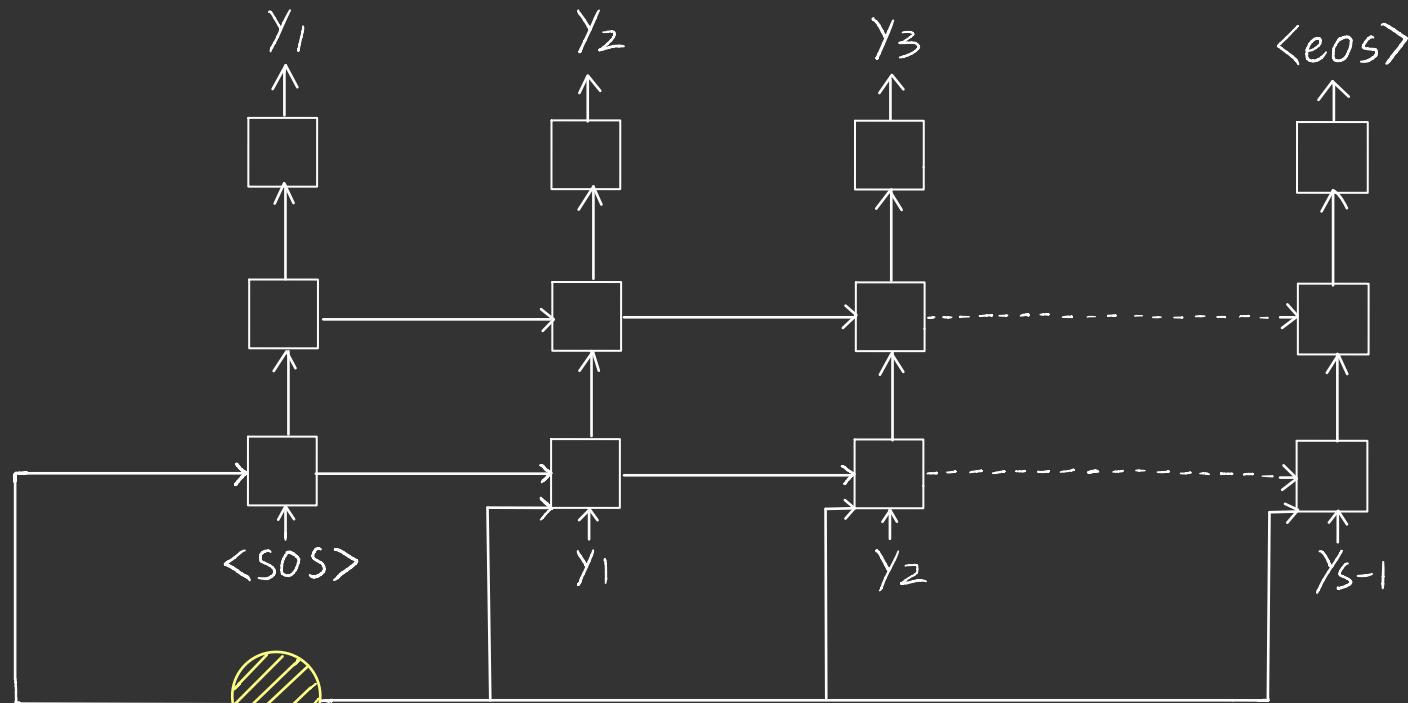


Vanilla Seq2Seq

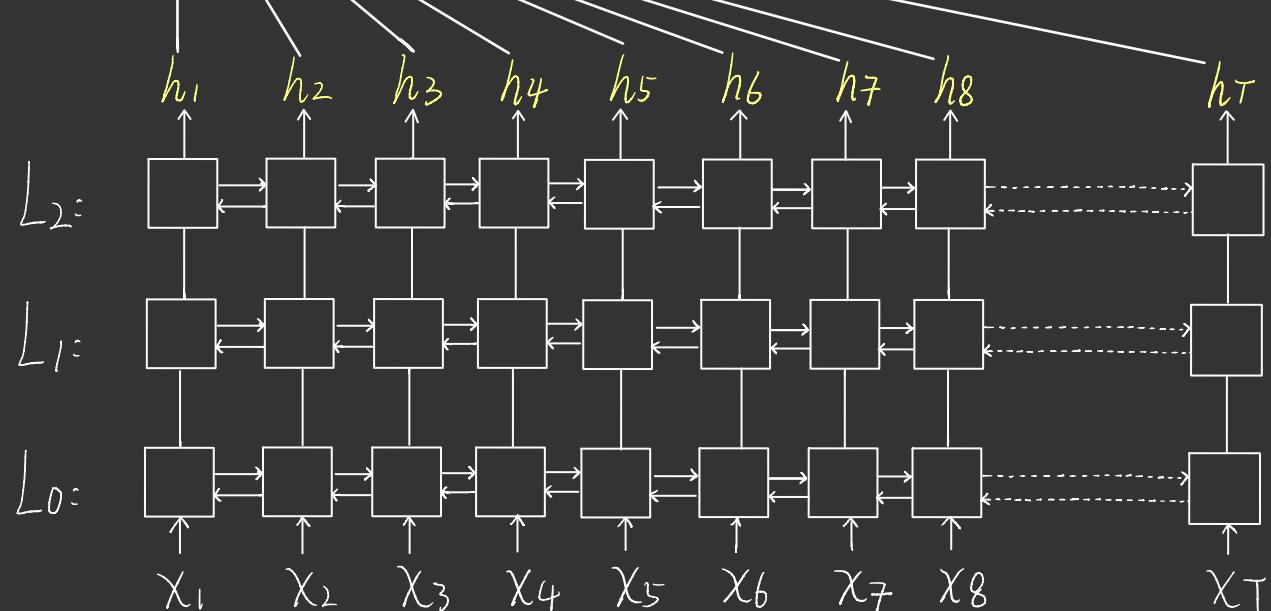


$$h = (h_1, \dots, h_T)$$

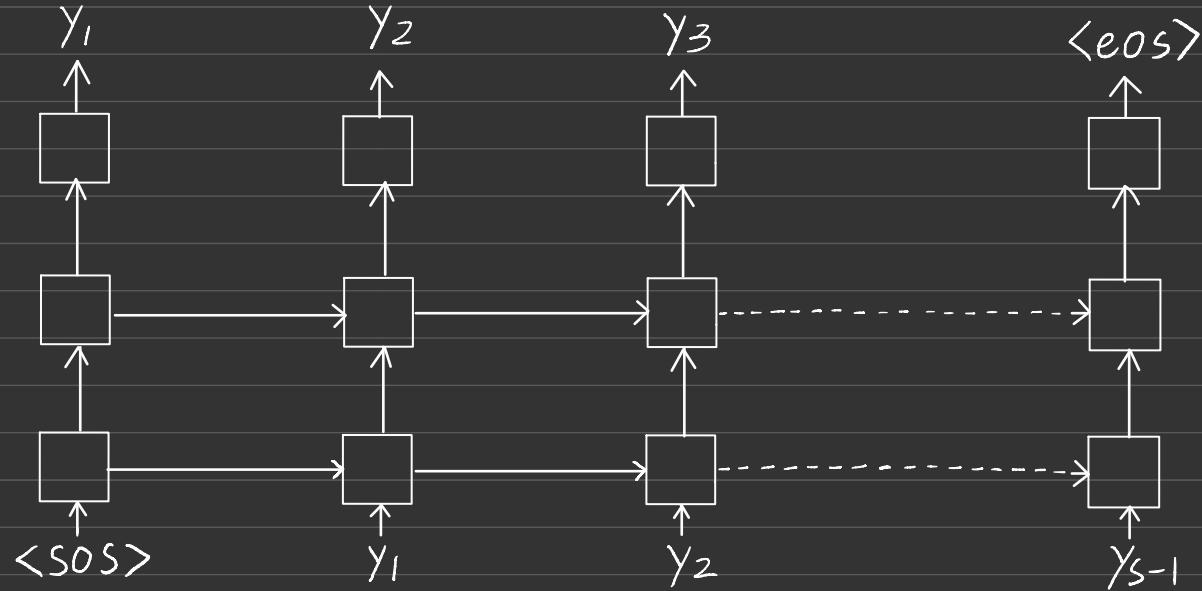




$$h = (h_1, \dots, h_U)$$



Attention



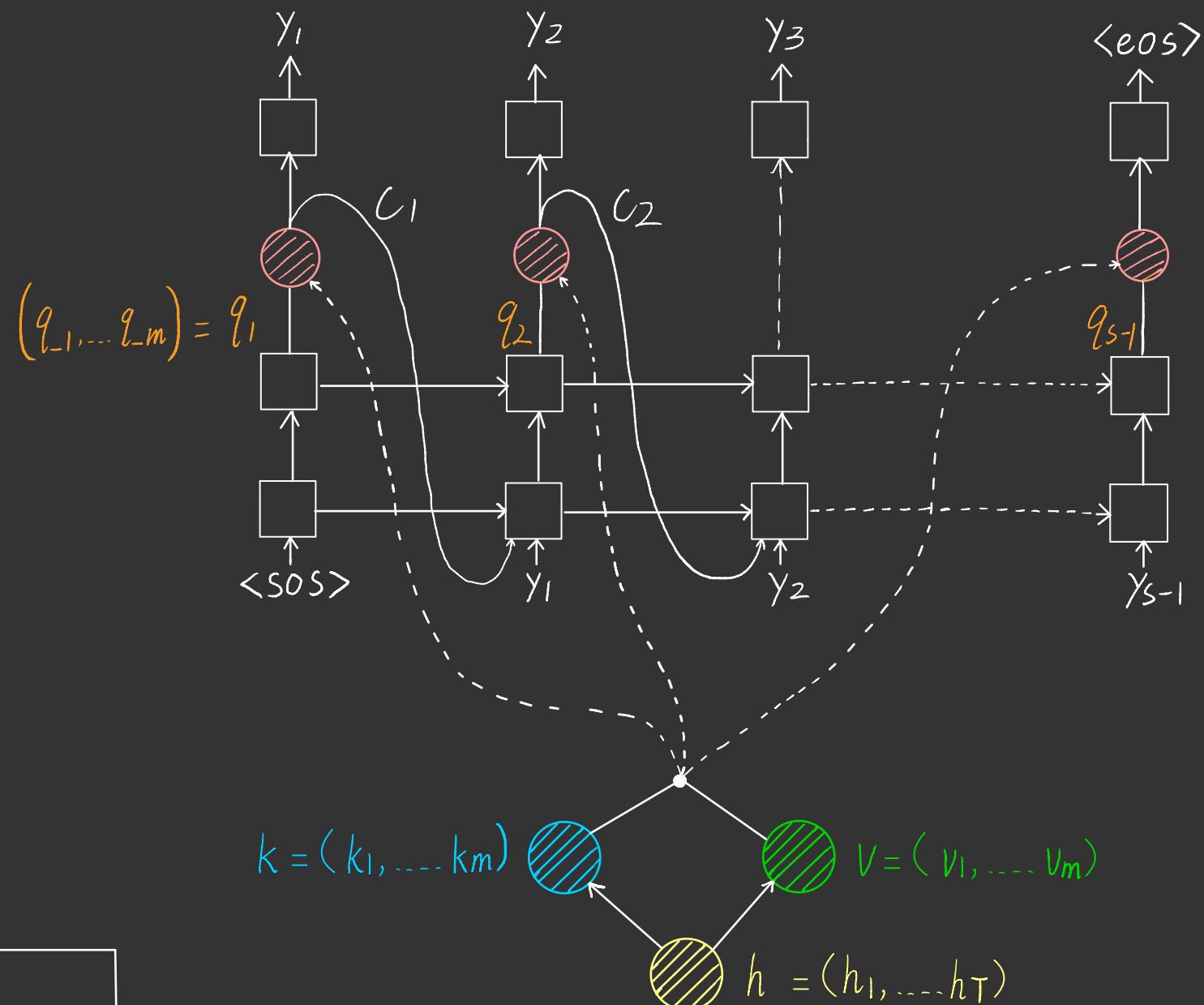
$h = (h_1, \dots, h_T)$

h_1	h_2	h_3	h_4	h_5	h_T
-------	-------	-------	-------	-------	-------	-------

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

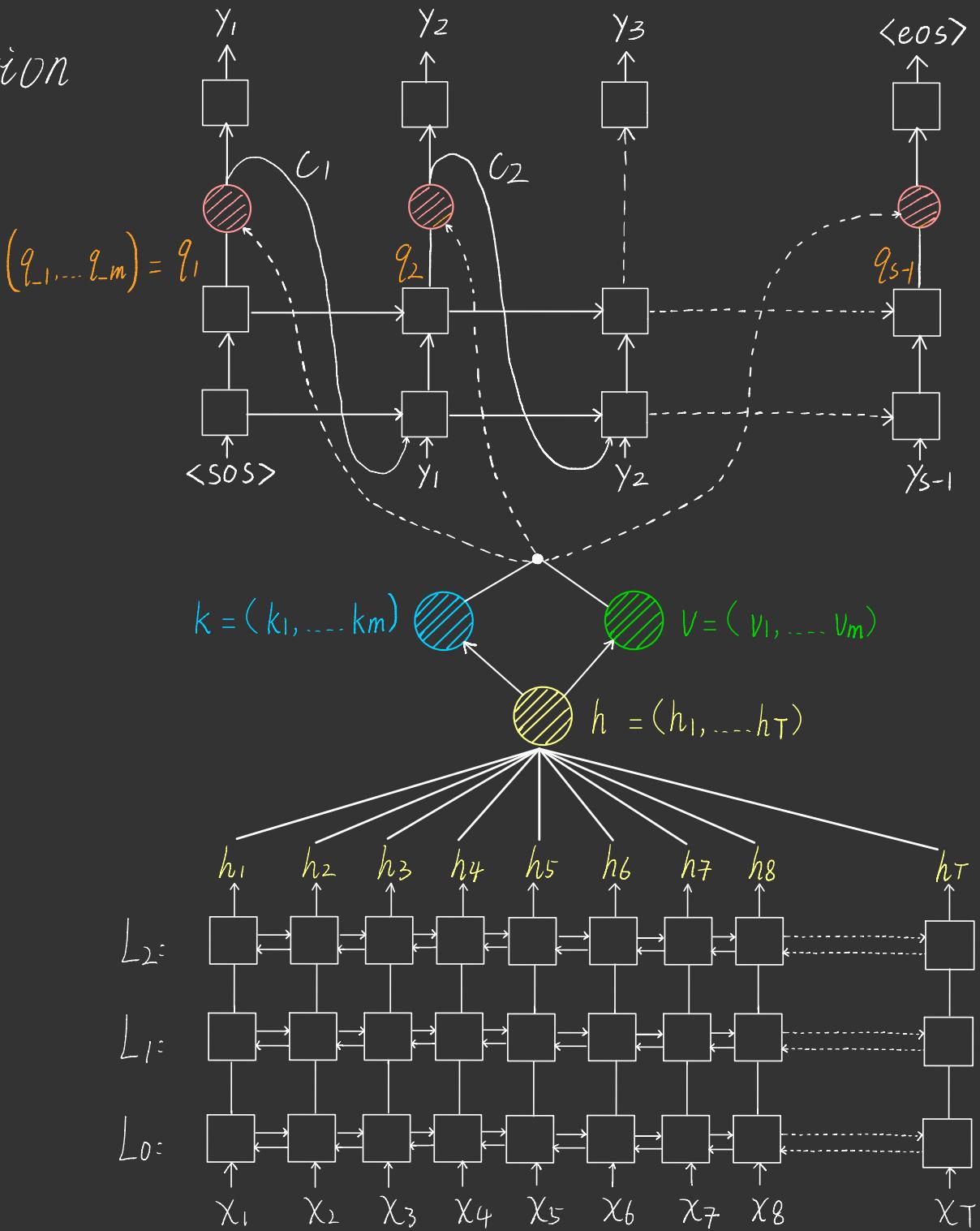
Eg. English: I love deep learning, it is interesting

Chinese: 我 爱 深度 学习, 它很 有趣

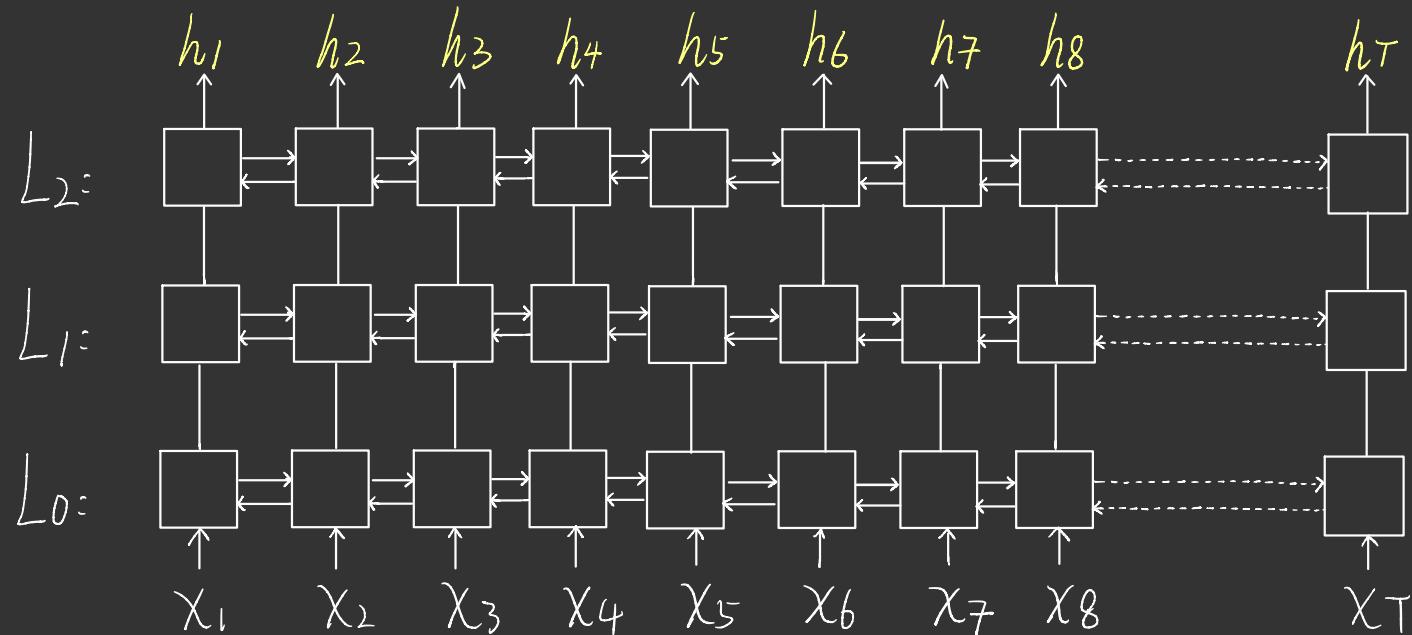


$C = \text{cos-sim}(q, k) \cdot V$

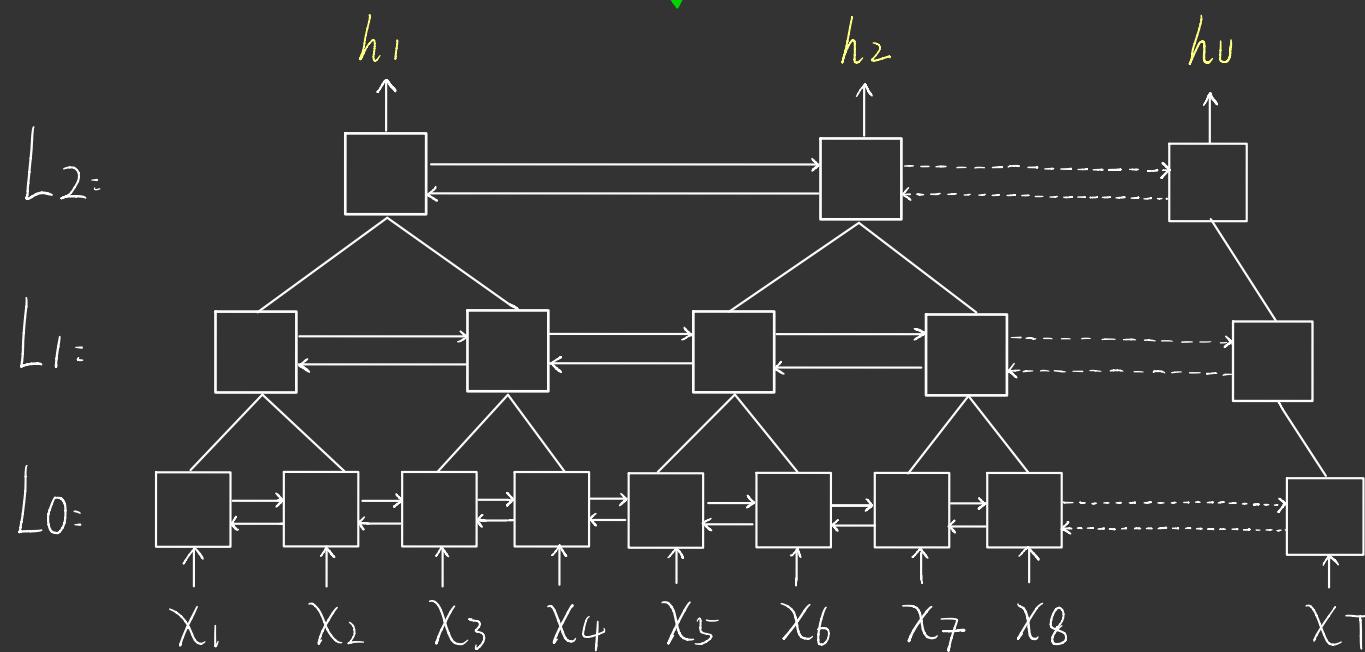
Seq2Seq + Attention



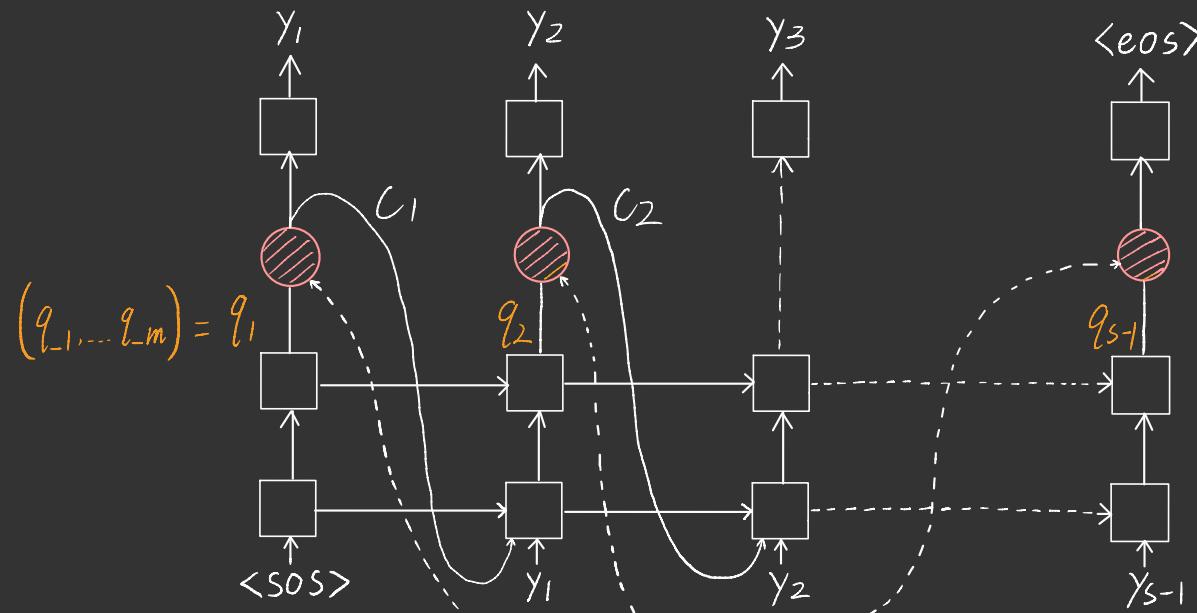
BLSTM



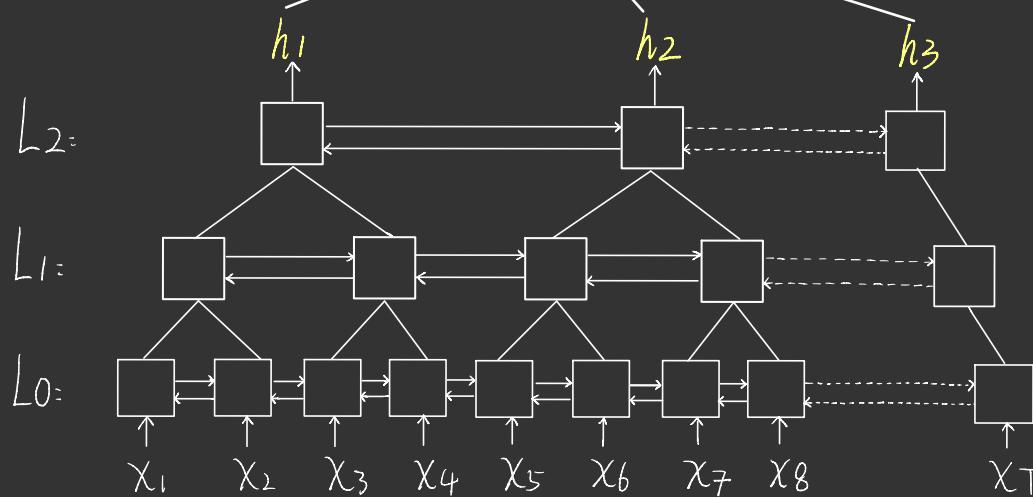
pBLSTM



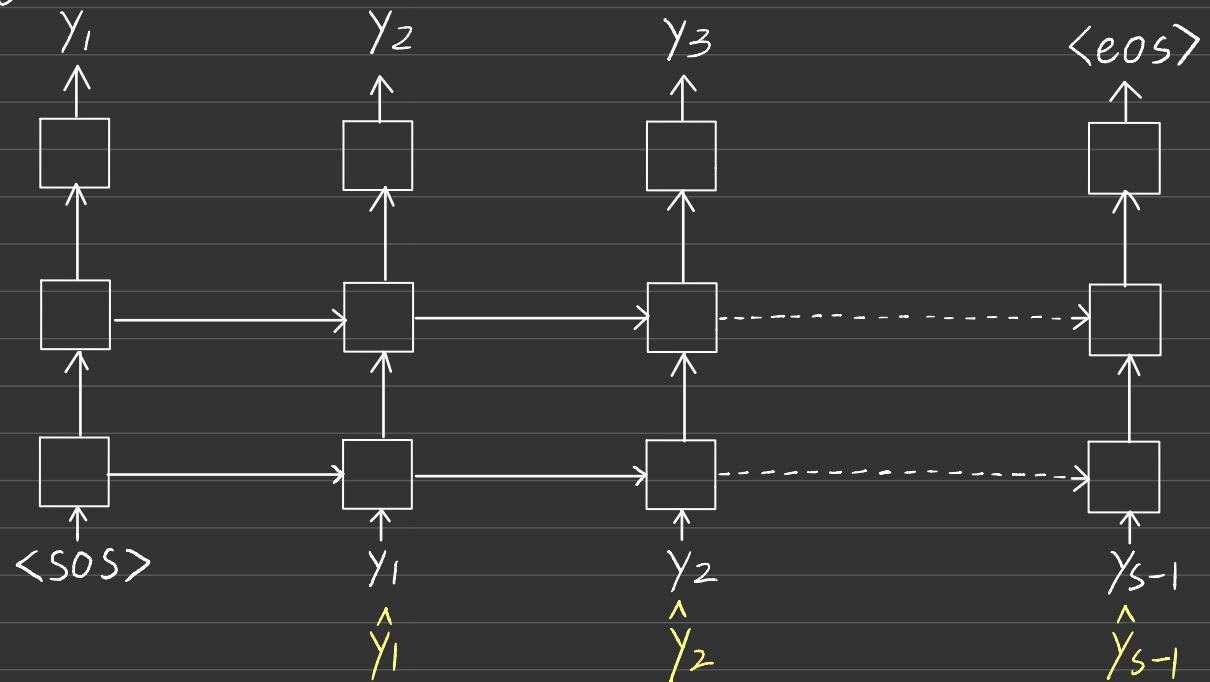
LAS



$$k = (k_1, \dots, k_m) \quad v = (v_1, \dots, v_m) \quad h = (h_1, \dots, h_T)$$



Teacher forcing:



Gumble noise

Data Augmentation

