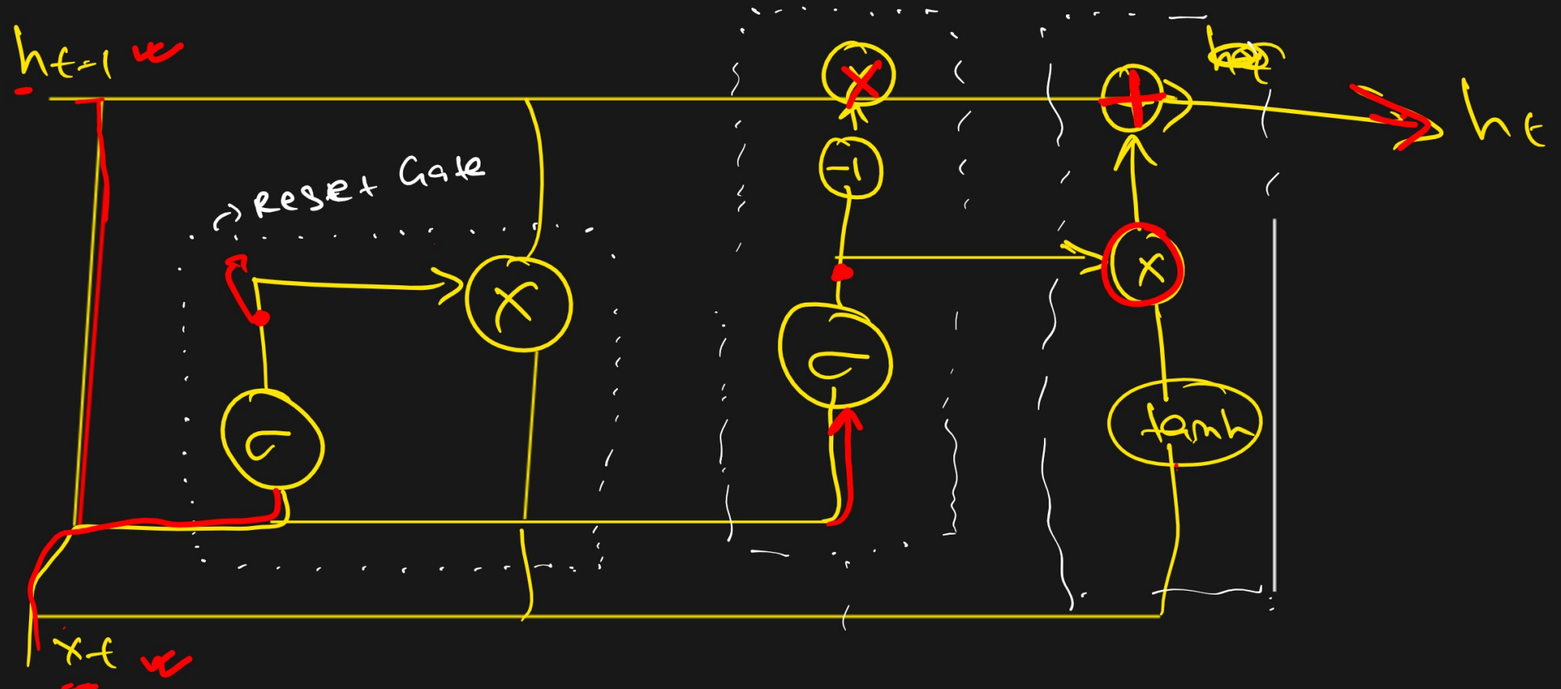
GRU → Gated Recurrent Unit

LSTM → ① Forget Gate
② Input Gate
③ Output Gate

GRU → ① Update Gate
② Reset Gate



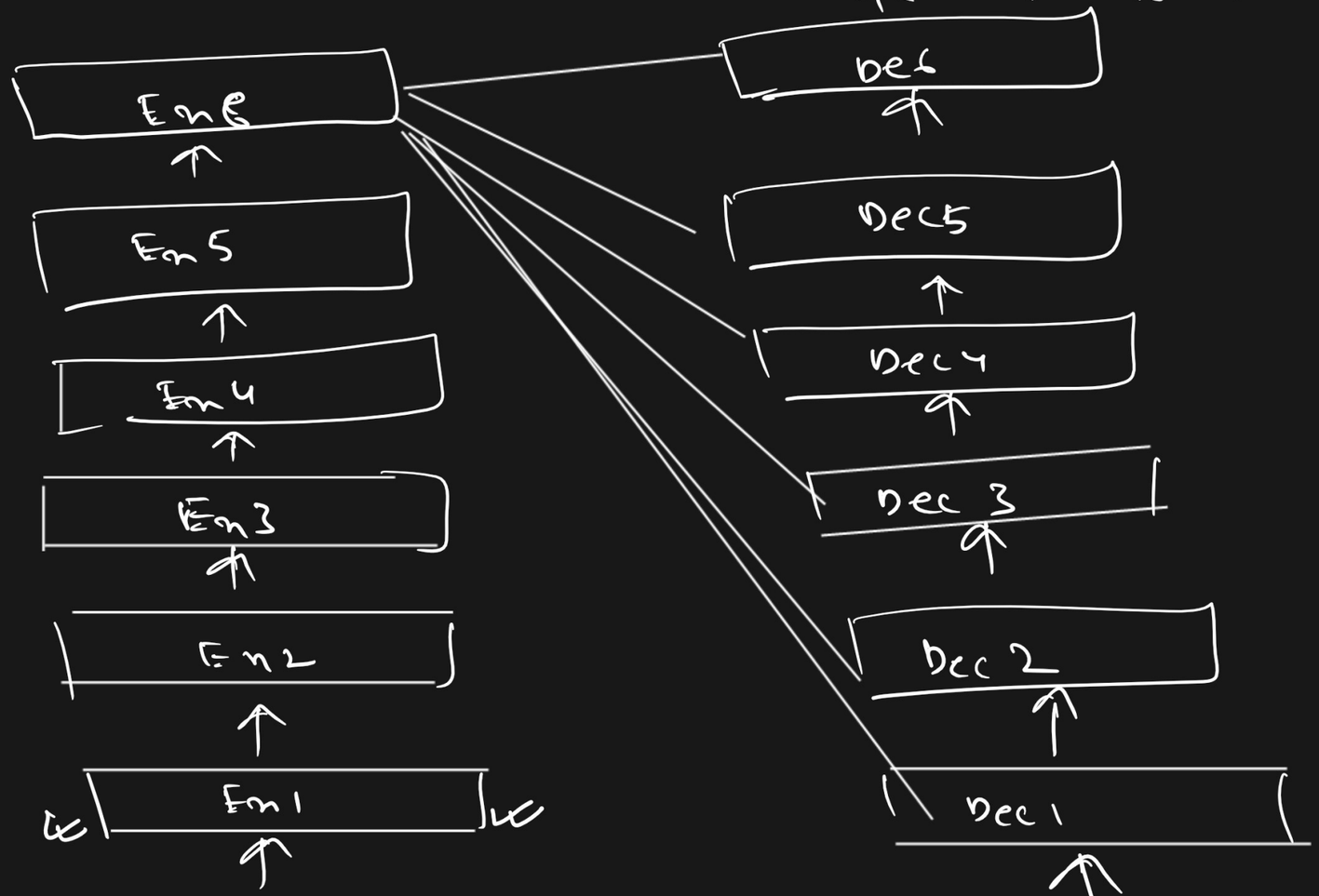
update gate

$$r_t = \sigma\left(w_h \cdot h_{t-1} + w_x \cdot x_t + b_r\right)$$

$$u_t = \sigma\left(w_{hu} \cdot h_{t-1} + w_{xu} \cdot x_t + b_u\right)$$

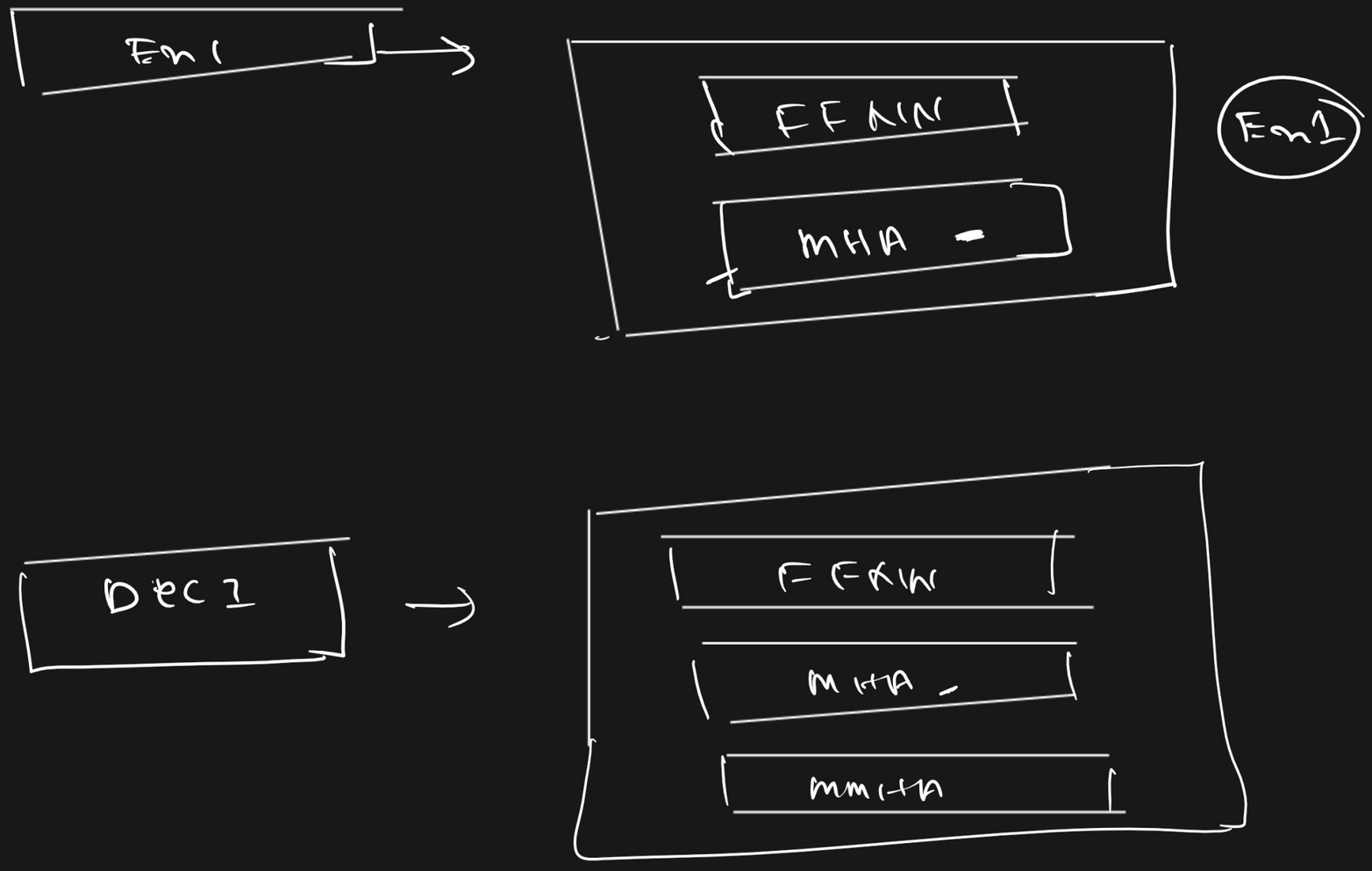
Attention is All you need

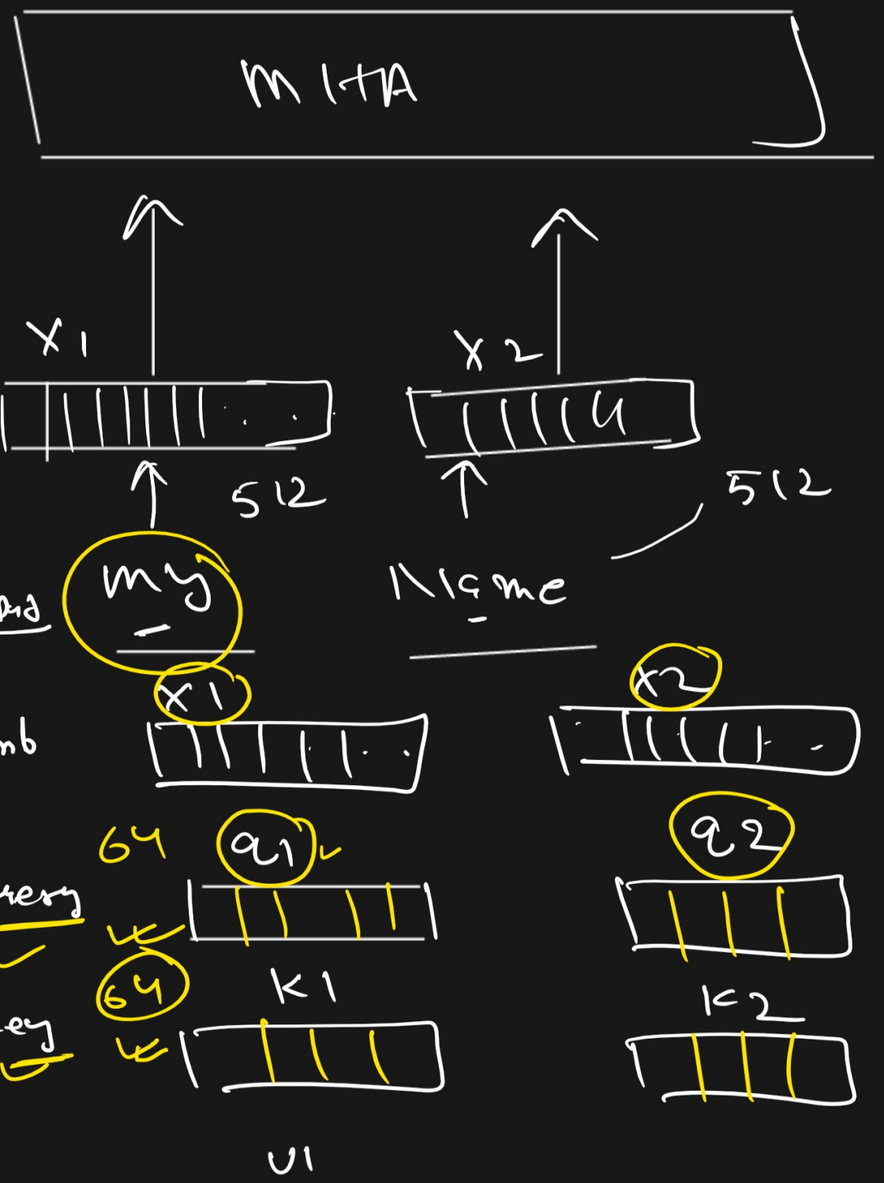
Encoder



Decoder

Input : My name is Sudhanshu Kumar

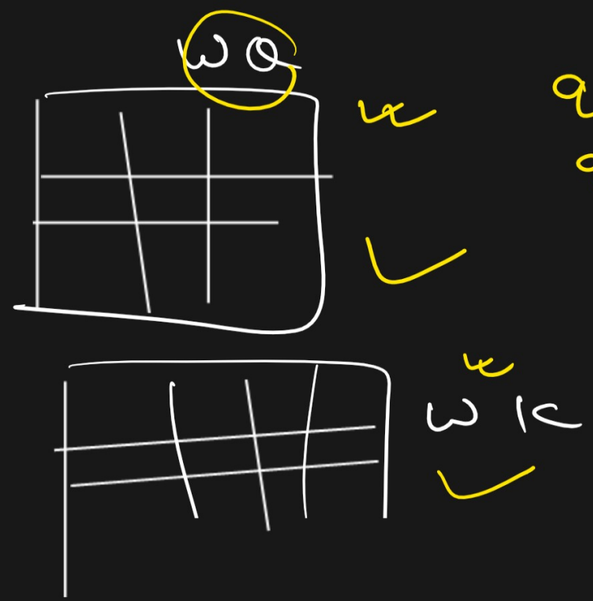

Enc 1
FFNN
MHA
Enc1
Dec 1
FFNN
MHA
MMHA



Q, K, U

given basic

Find basic



$q_1 = x_1 \cdot W_Q$

$q_2 = x_2 \cdot W_Q$

$K_1 = x_1 \cdot W_K$

$K_2 = x_2 \cdot W_K$

Value

v_2

w_v

$v_1 = x_1 \cdot w_v$

$v_2 = x_{12} \cdot w_v$

Alone

$q_2 k_1 =$

$q_2 k_2 =$

My

Score $q_1 \cdot k_1 = 100$ $q_1 k_2 = 90$

Divide $\left(\frac{q_1 k_1}{\sqrt{d_k}}\right)$ $\left(\frac{q_1 k_2}{\sqrt{d_k}}\right)$

Softmax $\left(\frac{100}{\sqrt{64}}\right)$ $\left(\frac{90}{\sqrt{64}}\right)$

$v_{11} =$

$v_{12} =$

Softmax x Value

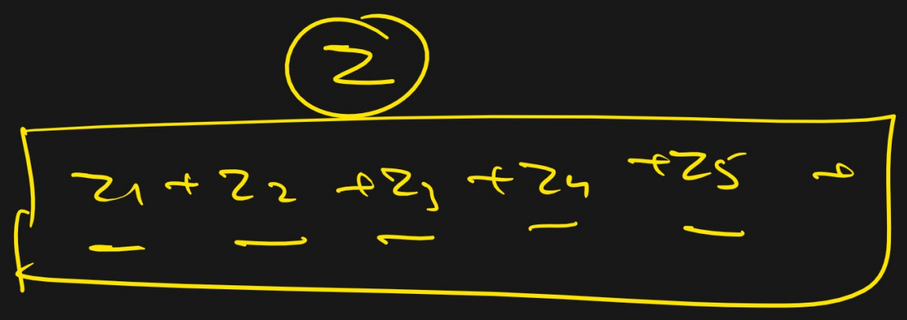
z_1 z_2 z_3 .....

Sum

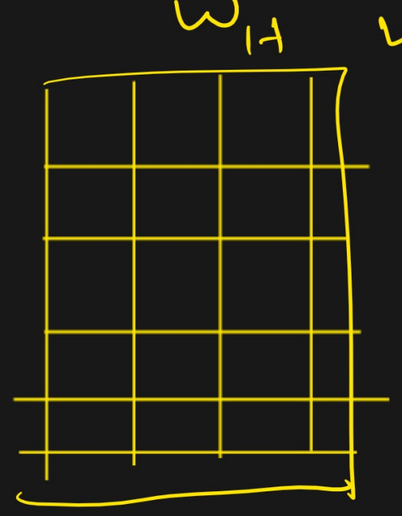
word → q → K → V → Score → Softmax → softmax value → add → vector

My Name is Sudhanshu Kumar

↓ ↓ ↓ ↓ ↓



X



My Name [illegible] Sudh , [illegible] teach [illegible] Science