

Figure 1: Scaled Dot Product Attention.

$$\underline{Q}^{3 \times 4} =) \quad (1a)$$

$$\underline{K}^{3 \times 4} =) \quad (1b)$$

$$\underline{V}^{3 \times 4} =) \quad (1c)$$

$$\underline{B}^{3 \times 4} = \text{mat_mult}(\underline{Q}^{3 \times 4}, \underline{K}^{3 \times 4}) \quad (1d)$$

$$\underline{Y}^{3 \times 4} = \text{scale}(\underline{B}^{3 \times 4}) \quad (1e)$$

$$\underline{R}^{3 \times 4} = \text{mask}(\underline{Y}^{3 \times 4}) \quad (1f)$$

$$\underline{G}^{3 \times 4} = \text{softmax}(\underline{R}^{3 \times 4}) \quad (1g)$$

$$P^{3 \times 4} = \text{mat_mult}(G^{3 \times 4}, V^{3 \times 4}) \quad (1h)$$

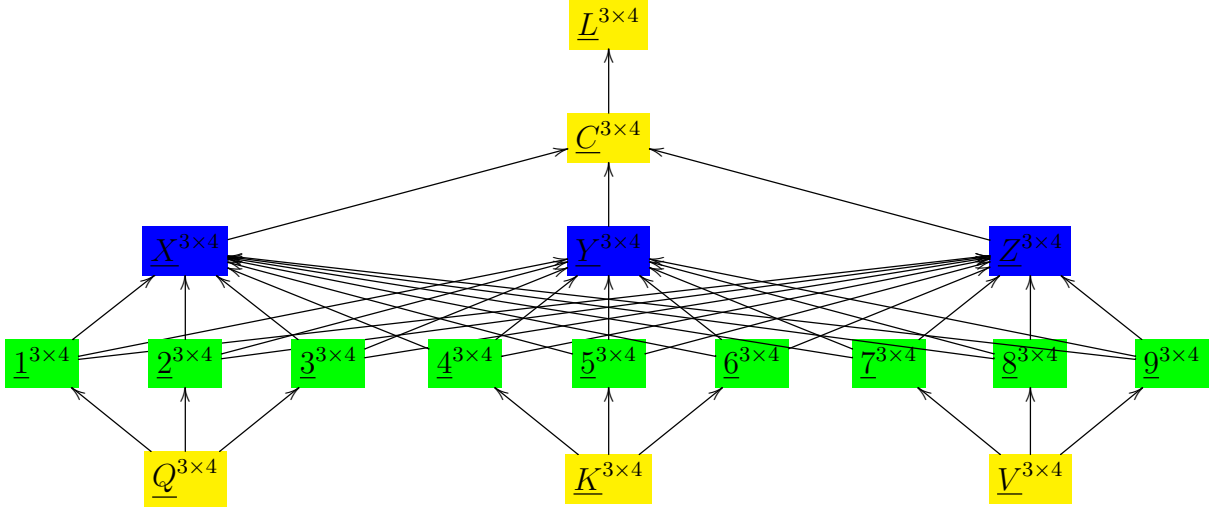


Figure 2: Multi-head Attention.

$$Q^{3 \times 4} =) \quad (2a)$$

$$K^{3 \times 4} =) \quad (2b)$$

$$V^{3 \times 4} =) \quad (2c)$$

$$\underline{1}^{3 \times 4} = \text{linear}(Q^{3 \times 4}) \quad (2d)$$

$$\underline{2}^{3 \times 4} = \text{linear}(Q^{3 \times 4}) \quad (2e)$$

$$\underline{3}^{3 \times 4} = \text{linear}(Q^{3 \times 4}) \quad (2f)$$

$$\underline{4}^{3 \times 4} = \text{linear}(K^{3 \times 4}) \quad (2g)$$

$$5^{3 \times 4} = \text{linear}(K^{3 \times 4}) \quad (2h)$$

$$6^{3 \times 4} = \text{linear}(K^{3 \times 4}) \quad (2i)$$

$$7^{3 \times 4} = \text{linear}(V^{3 \times 4}) \quad (2j)$$

$$8^{3 \times 4} = \text{linear}(V^{3 \times 4}) \quad (2k)$$

$$9^{3 \times 4} = \text{linear}(V^{3 \times 4}) \quad (2l)$$

$$X^{3 \times 4} = \text{scaled_dot_prod_att}(1^{3 \times 4}, 2^{3 \times 4}, 3^{3 \times 4}, 4^{3 \times 4}, 5^{3 \times 4}, 6^{3 \times 4}, 7^{3 \times 4}, 8^{3 \times 4}, 9^{3 \times 4}) \quad (2m)$$

$$Y^{3 \times 4} = \text{scaled_dot_prod_att}(1^{3 \times 4}, 2^{3 \times 4}, 3^{3 \times 4}, 4^{3 \times 4}, 5^{3 \times 4}, 6^{3 \times 4}, 7^{3 \times 4}, 8^{3 \times 4}, 9^{3 \times 4}) \quad (2n)$$

$$Z^{3 \times 4} = \text{scaled_dot_prod_att}(1^{3 \times 4}, 2^{3 \times 4}, 3^{3 \times 4}, 4^{3 \times 4}, 5^{3 \times 4}, 6^{3 \times 4}, 7^{3 \times 4}, 8^{3 \times 4}, 9^{3 \times 4}) \quad (2o)$$

$$C^{3 \times 4} = \text{concat}(X^{3 \times 4}, Y^{3 \times 4}, Z^{3 \times 4}) \quad (2p)$$

$$L^{3 \times 4} = \text{concat}(C^{3 \times 4}) \quad (2q)$$

$$q^{3 \times 4} = n^{3 \times 4}) \quad (3a)$$

$$k^{3 \times 4} = n^{3 \times 4}) \quad (3b)$$

$$n^{3 \times 4} = N^{3 \times 4}, F^{3 \times 4}) \quad (3c)$$

$$F^{3 \times 4} = N^{3 \times 4}) \quad (3d)$$

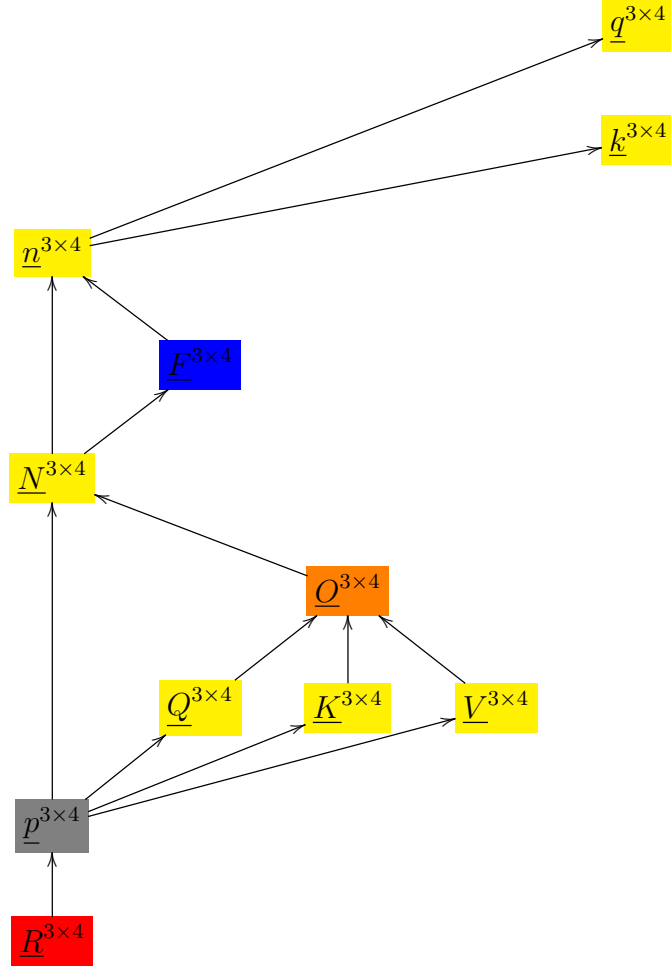


Figure 3: Encoder.

$$\underline{N}^{3 \times 4} = \underline{p}^{3 \times 4}, \underline{O}^{3 \times 4}) \quad (3e)$$

$$\underline{O}^{3 \times 4} = \underline{Q}^{3 \times 4}, \underline{K}^{3 \times 4}, \underline{V}^{3 \times 4}) \quad (3f)$$

$$\underline{Q}^{3 \times 4} = \underline{p}^{3 \times 4}) \quad (3g)$$

$$K^{3\times 4} = p^{3\times 4}) \tag{3h}$$

$$V^{3\times 4} = p^{3\times 4}) \tag{3i}$$

$$p^{3\times 4} = R^{3\times 4}) \tag{3j}$$

$$R^{3\times 4} =) \tag{3k}$$