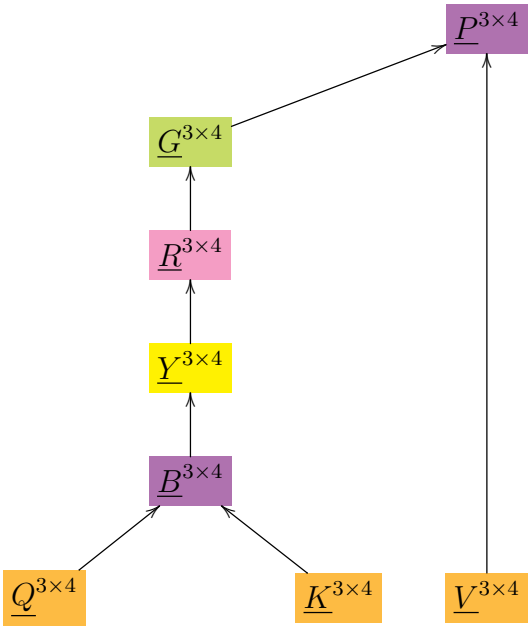

Figure 1: Scaled Dot Product Attention.

$$Q^{3\times4} =) \tag{1a}$$

$$K^{3\times4} =) \tag{1b}$$

$$V^{3\times4} =) \tag{1c}$$

$$B^{3\times4} = \text{mat\_mult}(Q^{3\times4}, K^{3\times4}) \tag{1d}$$

$$Y^{3\times4} = \text{scale}(B^{3\times4}) \tag{1e}$$

$$R^{3\times4} = \text{mask}(Y^{3\times4}) \tag{1f}$$

$$G^{3\times4} = \text{softmax}(R^{3\times4}) \tag{1g}$$

$$P^{3\times4} = \text{mat\_mult}(G^{3\times4}, V^{3\times4}) \tag{1h}$$