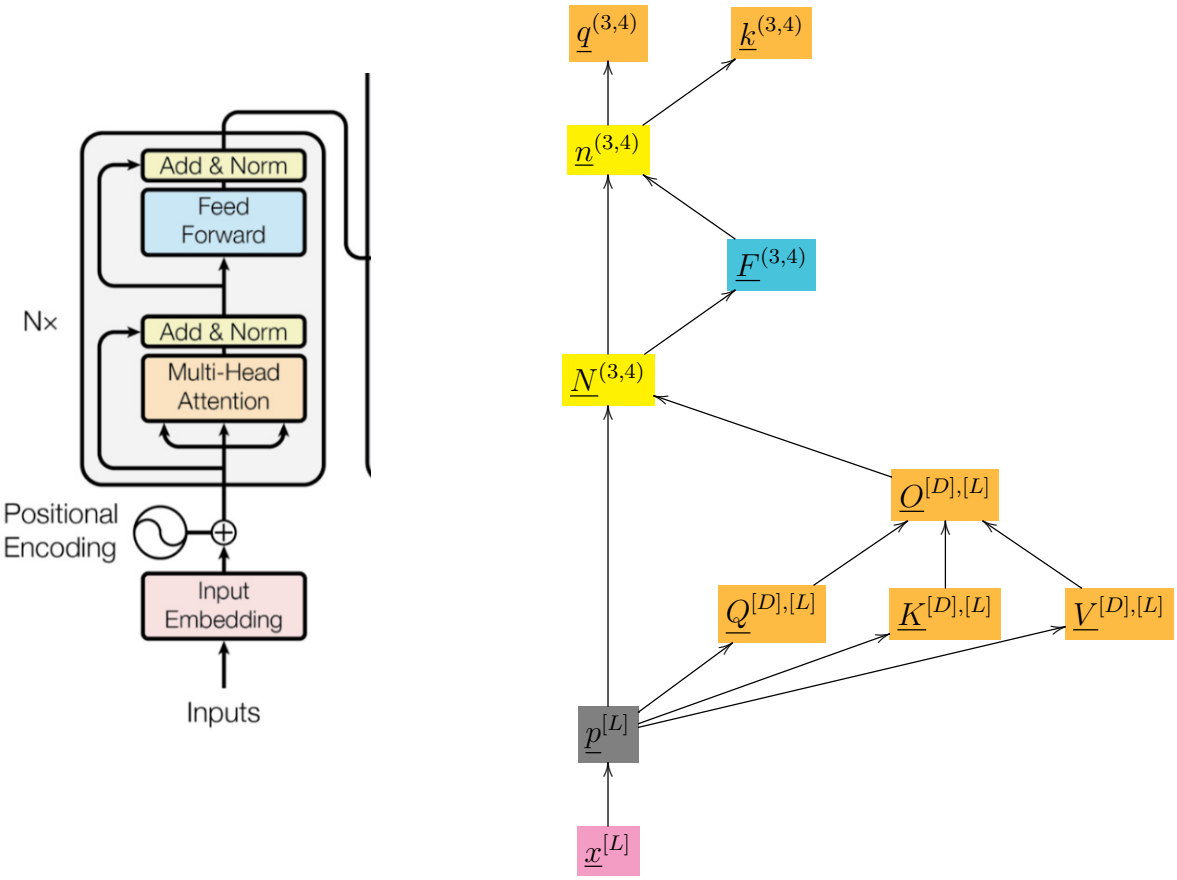

Figure 1: Encoder.

$$\underline{q}^{(3,4)} = \underline{n}^{(3,4)} \tag{1a}$$

$$\underline{k}^{(3,4)} = \underline{n}^{(3,4)} \tag{1b}$$

$$\underline{n}^{(3,4)} = \underline{N}^{(3,4)}, \underline{F}^{(3,4)} \tag{1c}$$

$$\underline{F}^{(3,4)} = \underline{N}^{(3,4)} \tag{1d}$$

$$\underline{N}^{(3,4)} = \underline{p}^{[L]}, \underline{O}^{[D],[L]} \tag{1e}$$

$$\underline{O}^{[D],[L]} = \text{multi\_headed\_attention}(\underline{Q}^{[D],[L]}, \underline{K}^{[D],[L]}, \underline{V}^{[D],[L]}) \tag{1f}$$

$$\underline{Q}^{[D],[L]} = W_{\underline{q}}^{[D],[d]} E^{[d],[L]} \tag{1g}$$

$$\underline{K}^{[D],[L]} = W_{\underline{k}}^{[D],[d]} E^{[d],[L]} \tag{1h}$$

$$\underline{V}^{[D],[L]} = W_{\underline{v}}^{[D],[d]} E^{[d],[L]} \tag{1i}$$

$$\underline{p}^{[L]} = M^{[L],[L]} \underline{x}^{[L]} \tag{1j}$$

$$\underline{x}^{[L]} = prior \tag{1k}$$