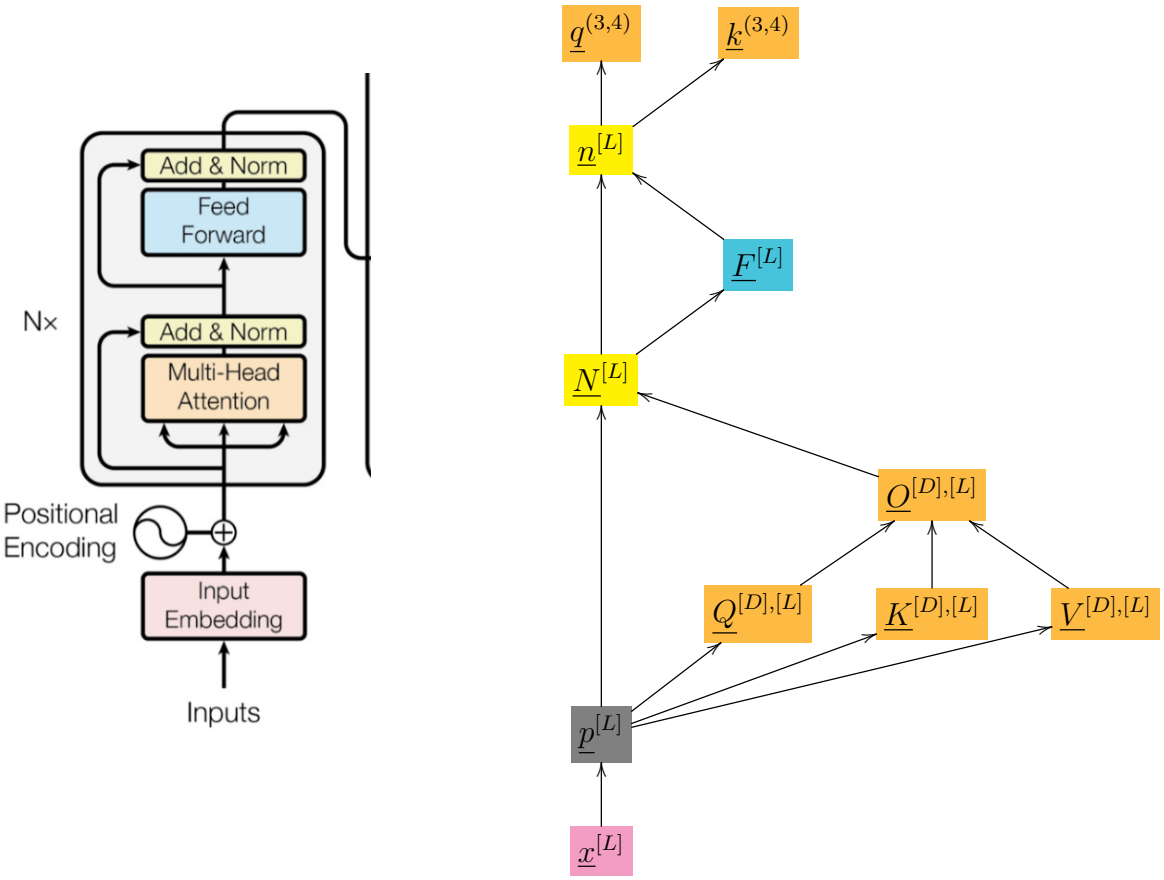

Figure 1: Encoder.

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

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

$$n^{[L]} = \text{normalize}(N^{[L]} + F^{[L]}) \tag{1c}$$

$$F^{[L]} = N^{[L]} \tag{1d}$$

$$N^{[L]} = \text{normalize}(p^{[L]} + O^{[D],[L]}) \tag{1e}$$

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

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

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

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

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

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