



Figure 1: Encoder.

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

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

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

$$\underline{F}^{[L]} = \underline{N}^{[L]} \quad (1d)$$

$$\underline{N}^{[L]} = \text{normalize}(\underline{p}^{[L]} + \underline{O}^{[L]}) \quad (1e)$$

$$O^{[L]} = \text{multi_headed_attention}(Q^{[D],[L]}, K^{[D],[L]}, V^{[D],[L]}) \quad (1f)$$

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

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

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

$$p^{[L]} = M^{[L],[L]} x^{[L]} \quad (1j)$$

$$x^{[L]} =; prior \quad (1k)$$