1.

$x = [1:l]$

$= x[1] \cdot x[2] \cdots x[l]$

$x[1] = length - de$

$x[1] \in R^{d_{in} \times 1} = R^{de \times 1}$

$d_{in} = de$

$x \in R^{de \times 1}$

$q[1] = W_q \cdot x[1] \in R^{d_{attn} \times 1}$

$R^{d_{attn} \times d_{in}} \quad R^{d_{in} \times 1}$

$x[1] \in R^{d_{in} \times 1}$, $W_q \in R^{d_{attn} \times d_{in}}$ $= W_k$

$k[1] = W_k \cdot x[1] = R^{d_{attn} \times d_{in}} \cdot R^{d_{in} \times 1}$

$= R^{d_{attn} \times 1}$

$V[1] \in R^{d_{out}} = R^{d_{out} \times d_{in}} \cdot R^{d_{in} \times 1}$

$= W_v \cdot x[1]$

$W_v = R^{d_{out} \times d_{in}}$

$\in R^{1 \times d_{hidden}}$

$W_2 \quad W_1 \tilde{V}[1] = R^{d_{hidden} \times d_{out}} \times R^{d_{out} \times l}$

$= \tilde{x}[1]$

$q[1]^T \in R^{1 \times d_{attn}}$

$k[1] \in R^{d_{attn} \times 1}$

$\in R^{1 \times 1}$

$W_2 \in R^{d_{hidden} \times 1}$

$W_1 \in R^{d_{hidden} \times d_{out}}$

$a[1,1] \quad V[1]$

$R^{1 \times 1} \times d_{out} \in R^{d_{out} \times 1}$

$q[1]^T \cdot k[1] \in R^{1 \times 1}$

$\sum_{i=1}^{l} q[i]^T \cdot k[i] \in R^{1 \times l}$

$d_{out} = l$ <1>

$\tilde{V}[1] = \sum_{j=1}^{l} a[1,j] \cdot V[j]$

$\in R^{1 \times l} \cdot R^{d_{out} \times 1}$

$\tilde{V}[1] \in R^{1 \times l}$

$W_1 \cdot \tilde{V}[1] = R^{d_{hidden} \times d_{out}} \cdot R^{l \times l}$

$= R^{d_{hidden} \times l}$

$W_2^T \in R^{1 \times d_{hidden}}$

$y[1] \in R^{1 \times l}$

$W_q \in R^{d_{attn} \times d_{in}}$, $W_K \in R^{d_{attn} \times d_{in}}$

$W_V = R^{d_{out} \times d_{in}}$, $W_1 \in R^{d_{hidden} \times d_{out}}$,

$W_2 \in R^{d_{hidden} \times 1}$, $x \in R^{de \times l} = R^{d_{in} \times l}$

$\lceil d_{in} = de,\ l = d_{out} \rfloor$

$W_1 \in R^{2 \times d_{out}} = d_{hidden} \times d_{out}$.

$W_V \in R^{d_{out} \times d_{in}} = R^{d_{out} \times d_{out}}$

$= R^{l \times l}$

$de = l = d_{out} = d_{in}$, $d_{hidden} = 2$

$W_q = d_{attn} \times l$, $W_K = d_{attn} \times l$

$W_V = l \times l$, $W_1 = 2 \times l$

$W_2 = 2 \times 1$

$\rightarrow l^2 + 2l + 2l \cdot d_{attn} + 2$

$= \boxed{l^2 + 2l \cdot (1 + d_{attn}) + 2}$

$\hat{y}[i]$

$= W_2^T \cdot \mathrm{ReLU}(W_1 \cdot \tilde{v}[i])$

$W_1 \in R^{2 \times d_{out}} = R^{2 \times l}$.

$\tilde{v}[i] \in R^{l \times l} = R^{d_{out} \times l}$

$W_1 \tilde{v}[i] \in R^{2 \times l} \cdot R^{l \times l}$

$= \boxed{4l^2}$

$W_2^T \in R^{1 \times 2} \cdot (W_1 \tilde{v}[i] \in R^{2 \times l})$

$= \boxed{4l}$.

$\therefore 4l + 4l^2$

$= 4l(1 + l)$

: feed-forward block

3. activation $\tilde{v}[1]\ \tilde{v}[2] \cdots \tilde{v}[l]$

$\tilde{v}[i] \in R^{l \times l}$ — $i = 1, 2, \cdots, l$

↓

$l \times l^2$ = size <3>

$R^{l \times l} \cdot R^{d_{out} \times 1} = R^{l \times l} \cdot R^{l \times 1}$

$\Rightarrow \boxed{2l^2}$

↓

FLOP = $2d_{atten} + 2d_{atten} \cdot l + 2l^2$

2. $a[i, j] = \dfrac{① \exp(q[i]^T \cdot k[j] / \sqrt{d_{atten}})}{② \sum_{u=1}^{l} \exp(q[i]^T \cdot k[u] / \sqrt{d_{atten}})}$

$\Rightarrow \boxed{2d_{atten} \cdot (1+l) + 2l^2}$

$q^T[i] \in R^{1 \times d_{atten}}$, $k[j] \in R^{d_{atten} \times 1}$

$q^T[i] \cdot k[j] \in R^{1 \times 1}$

$\sum_{u=1}^{l} \exp(q^T[i] \cdot k[u] / \sqrt{d_{atten}}) \in R^{1 \times l}$

① $q[i]^T \cdot k[j] \in R^{1 \times 1}$

$\boxed{2 d_{atten}}$

② $R^{1 \times d_{atten}} \cdot R^{d_{atten} \times l} = \boxed{2 d_{atten} \cdot l}$

$\tilde{v}[i] = \sum_{j=1}^{l} a[i,j] \cdot v[j]$

$R^{l \times l} \cdot \in R^{d_{out} \times 1}$

6. $l(y - \hat{y}) = \sum_{i=1}^{l} (y[i] - \hat{y}[i])^2$

$\hat{y}[i] \in R^{1 \times l}$

$y^2[i] - 2y[i]\hat{y}[i] + \hat{y}^2[i]$

$\hat{y}[i]$의 FLOPs $= 4l^2$

$w_2^T \in R^{1 \times 2} \cdot w_1 \in R^{2 \times d_{out}} \cdot \tilde{v}[i] \in R^{l \times 1}$

$= w_2^T \in R^{1 \times 2} \cdot R^{2 \times l}$

$4l^2 + 4l \times 3$

$\Rightarrow \boxed{12l(l+1)}$: $y^2[i]$, $-2y[i] \cdot \hat{y}[i]$, $\hat{y}^2[i]$ FLOPs 등의 합.

4. backward propagation

$W_2^T \in R^{1 \times d_{hidden}} = R^{1 \times 2}$

$W_1 \in R^{d_{hidden} \times d_{out}}$

$= R^{2 \times l}$ <4>

$W_2^T W_1 \in R^{1 \times l}$ $\lceil 4l \rfloor$

$W_2^T W_1 \cdot \tilde{v}[i] = R^{1 \times l} \cdot R^{l \times l}$

$= R^{1 \times l}$ $\lceil 2l^2 \rfloor$

$\therefore$ $4l + 2l^2 = 2l(2+l)$

5. $W_q \in R^{d_{attn} \times d_{in}} = R^{d_{attn} \times l}$

$$= \begin{bmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 0 & 2M^2 & 0 & \cdots & 0 & & \\ \vdots & & & & & & \\ 0 & \cdots & \cdots & 0 & 2M^2 & 0 & \cdots 0 \end{bmatrix}$$

$W_q(t,t) = 2M^2$, $t = 2, \cdots, d_{attn}$

$W_K \in R^{d_{attn} \times l}$

$$= \begin{bmatrix} 0 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & & \vdots \\ 0 & 0 & 1 & \cdots & 0 & \vdots \\ \vdots & & & \ddots & & \vdots \\ 0 & \cdots & & 0 & 1 & 0 \cdots 0 \end{bmatrix}$$

$W_K(t,t) = 1$, $t = 2, 3, \cdots, d_{attn}$

$W_{V,(i,j)} \in R^{l \times l}$ $\begin{cases} 1, & i = n+4,\ j = l \\ 0, & \text{elsewhere} \end{cases}$